

# Reduced-Dimension Linear Transform Coding of Correlated Signals in Networks

Naveen Goela, *Student Member, IEEE*, and Michael Gastpar<sup>†</sup>, *Member, IEEE*

## Abstract

A model, called the linear transform network (LTN), is proposed to analyze the compression and estimation of correlated signals transmitted over directed acyclic graphs (DAGs). An LTN is a DAG network with multiple source and receiver nodes. Source nodes transmit subspace projections of random correlated signals by applying reduced-dimension linear transforms. The subspace projections are linearly processed by multiple relays and routed to intended receivers. Each receiver applies a linear estimator to approximate a subset of the sources with minimum mean squared error (MSE) distortion. The model is extended to include noisy networks with power constraints on transmitters. A key task is to compute all local compression matrices and linear estimators in the network to minimize end-to-end distortion. The non-convex problem is solved iteratively within an optimization framework using constrained quadratic programs (QPs). The proposed algorithm recovers as special cases the regular and distributed Karhunen-Loève transforms (KLTs). Cut-set lower bounds on the distortion region of multi-source, multi-receiver networks are given for linear coding based on convex relaxations. Cut-set lower bounds are also given for any coding strategy based on information theory. The distortion region and compression-estimation tradeoffs are illustrated for different communication demands (e.g. multiple unicast), and graph structures.

This work was supported in part by the National Science Foundation under Grant CCF-0627024 and made with U.S. Government support under and awarded by DoD, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a. The material in this paper was presented in part at the IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan in April 2009 and at the IEEE International Symposium on Information Theory, Seoul, South Korea in June 2009.

Copyright © 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

N. Goela and M. Gastpar are with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, CA 94720-1770 USA (e-mail: {ngoela, gastpar}@eecs.berkeley.edu).

<sup>†</sup>M. Gastpar is also with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale (EPFL), Lausanne, Switzerland.

### Index Terms

Karhunen-Loève transform (KLT), linear transform network (LTN), quadratic program (QP), cut-set bound.

## I. INTRODUCTION

**T**HE compression and estimation of an observed signal via subspace projections is both a classical and current topic in signal processing and communication. While random subspace projections have received considerable attention in the compressed sensing literature [1], subspace projections optimized for minimal distortion are important for many applications. The Karhunen-Loève transform (KLT) and its empirical form Principal Components Analysis (PCA), are widely studied in computer vision, biology, signal processing, and information theory. Reduced dimensionality representations are useful for source coding, noise filtering, compression, clustering, and data mining. Specific examples include eigenfaces for face recognition, orthogonal decomposition in transform coding, and sparse PCA for gene analysis [2]–[4].

In contemporary applications such as wireless sensor networks (WSNs) and distributed databases, data is available and collected in different locations. In a WSN, sensors are usually constrained by limited power and bandwidth resources. This has motivated existing approaches to take into account correlations across high-dimensional sensor data to reduce transmission requirements (see e.g. [5]–[11]). Rather than transmitting raw sensor data to a fusion center to approximate a global signal, sensor nodes carry out local data dimensionality reduction to increase bandwidth and energy efficiency.

In the present paper, we propose a linear transform network (LTN) model to analyze dimensionality reduction for compression-estimation of correlated signals in *multi-hop* networks. In a centralized setting, given a random source signal  $\mathbf{x}$  with zero-mean and covariance matrix  $\Sigma_{\mathbf{x}}$ , applying the KLT to  $\mathbf{x}$  yields uncorrelated components in the eigenvector basis of  $\Sigma_{\mathbf{x}}$ . The optimal linear least squares  $k^{th}$ -order approximation of the source is given by the  $k$  components corresponding to the  $k$  largest eigenvalues of  $\Sigma_{\mathbf{x}}$ . In a network setting, multiple correlated signals are observed by different source nodes. The source nodes transmit low-dimensional subspace projections (approximations of the source) to intended receivers via a relay network. The compression-estimation problem is to optimize the subspace projections computed by all nodes in order to minimize the end-to-end distortion at receiver nodes.

In our model, receivers estimate random vectors based on “one-shot” linear *analog-amplitude* multisensor observations. The restriction to “one-shot”, zero-delay encoding of each vector of source observations separately is interesting due to severe complexity limitations in many applications (e.g. sensor networks).

Linear coding depends on first-order and second-order statistics and is robust to uncertainty in the precise probabilistic distribution of the sources. Under the assumption of ideal channels between nodes, our task is to optimize signal subspaces given limited bandwidth in terms of the number of real-valued messages communicated. Our results extend previous work on distributed estimation in this case [5]–[8]. For the case of dimensionality-reduction with noisy channel communication (see e.g. [6]), the task is to optimize signal subspaces subject to channel noise and power constraints.

For noisy networks, the general communication problem is often referred to as the *joint source-channel-network coding problem* in the information-theoretic literature and is a famously open problem. Beyond the zero-delay, linear dimensionality-reduction considered here, end-to-end performance in networks could be improved by (i), non-linear strategies and (ii), allowing a longer coding horizon. Partial progress includes non-linear low-delay mappings for only simple network scenarios [12]–[14]. For the case of an infinite coding horizon, separation theorems for decomposing the joint communication problem have been analyzed by [15]–[17].

#### A. Related Work

Directly related to our work in networks is the *distributed KLT* problem. Distributed linear transforms were introduced by Gastpar et al. for the compression of jointly Gaussian sources using iterative methods [5] [18]. Simultaneous work by Zhang et al. for multi-sensor data fusion also resulted in iterative procedures [8]. An alternate proof based on innovations for second order random variables with arbitrary distributions was given by [19]. The problem was extended for non-Gaussian sources, including channel fading and noise effects to model the non-ideal link from sensors to decoder by Schizas et al. [6]. Roy and Vetterli provide an *asymptotic* distortion analysis of the distributed KLT, in the case when the dimension of the source and observation vectors approaches infinity [20]. Finally, Xiao et al. analyze linear transforms for distributed *coherent* estimation [7].

Much of the estimation-theoretic literature deals with *single-hop* networks; each sensor relays information directly to a fusion center. In *multi-hop* networks, linear operations are performed by successive relays to aggregate, compress, and redistribute correlated signals. The LTN model relates to recent work on routing and *network coding* (Ahlsweide et al. [21]). In pure routing solutions, intermediate nodes either forward or drop packets. The corresponding analogy in the LTN model is to constrain transforms to be essentially identity transforms. However, network coding (over finite fields) has shown that mixing of data at intermediate nodes achieves higher rates in the multicast setting (see [22] regarding the sufficiency of linear codes and [23] for multicast code construction). Similarly in the LTN model, linear combining of subspace projections (over

the real field) at intermediate nodes improves decoding performance. Lastly, the max-flow min-cut theorem of Ford-Fulkerson [24] provides the basis for cut-set lower bounds in networks.

The LTN model is partially related to the formulation of Koetter and Kschischang [25] modeling information transmission as the injection of a basis for a vector space into the network, and subspace codes [26]. If arbitrary data exchange is permitted between network nodes, the compression-estimation problem is related to estimation in graphical models (e.g. decomposable PCA [27], and tree-based transforms (tree-KLT) [28]). Other related work involving signal projections in networks includes joint source-channel communication in sensor networks [29], random projections in a gossip framework [30], and distributed compressed sensing [31].

### B. Summary of Main Results

We cast the network compression-estimation problem as a statistical signal processing and constrained optimization problem. For most networks, the optimization is non-convex. Therefore, our main results are divided into two categories: (i) Iterative solutions for linear transform coding over acyclic networks; (ii) Cut-set bounds based on convex relaxations and cut-set bounds based on information theory.

- Section III reviews linear signal processing in networks. Section IV outlines an iterative optimization for compression-estimation matrices in ideal networks under a local convergence criterion.
- Section V analyzes an iterative optimization method involving constrained quadratic programs for noisy networks with power allocation over subspaces.
- Section VI introduces cut-set lower bounds to benchmark the minimum mean square error (MSE) for linear coding based on convex relaxations such as a semi-definite program (SDP) relaxation.
- Section VI-F describes cut-set lower bounds for any coding strategy in networks based on information-theoretic principles of source-channel separation. The lower bounds are plotted for a distributed noisy network.
- Sections IV-VI provide examples illustrating the tradeoffs between compression and estimation; upper and lower bounds are illustrated for an aggregation (tree) network, butterfly network, and distributed noisy network.

### C. Notation

Boldface upper case letters denote matrices, boldface lower case letters denote column vectors, and calligraphic upper case letters denote sets. The  $\ell^2$ -norm of a vector  $\mathbf{x} \in \mathbb{R}^n$  is defined as  $\|\mathbf{x}\|_2 \triangleq \sqrt{\sum_{i=1}^n |x_i|^2}$ . The weighted  $\ell^2$ -norm  $\|\mathbf{x}\|_{\mathbf{W}} \triangleq \|\mathbf{W}\mathbf{x}\|_2$  where  $\mathbf{W}$  is a positive semi-definite matrix (written  $\mathbf{W} \succeq \mathbf{0}$ ).

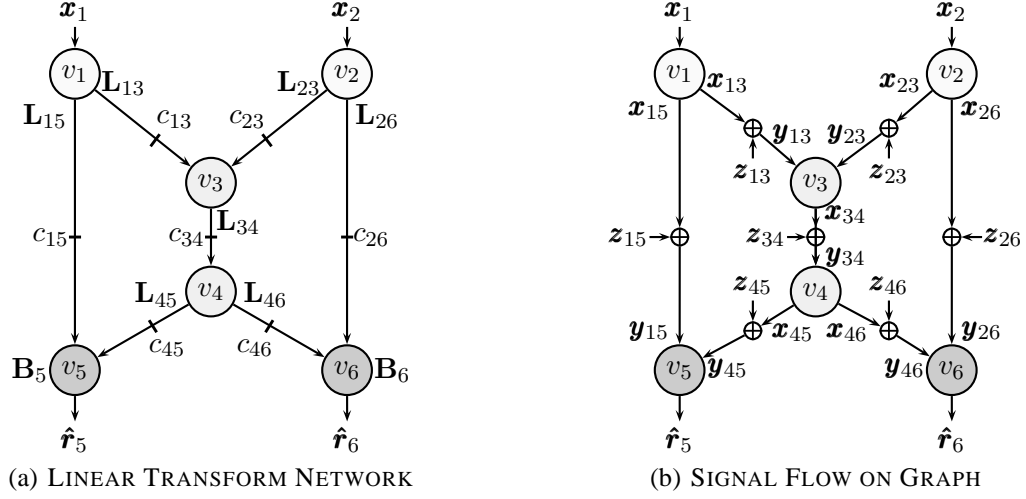


Fig. 1. (a) *Linear Transform Network*: An LTN model with source nodes  $\{v_1, v_2\}$  and receivers  $\{v_5, v_6\}$ . Source nodes observe vector signals  $\{\mathbf{x}_1, \mathbf{x}_2\}$ . All encoding nodes linearly process received signals using a transform  $\mathbf{L}_{ij}$ . Receivers  $v_5$  and  $v_6$  compute LLSE estimates  $\hat{\mathbf{r}}_5$  and  $\hat{\mathbf{r}}_6$  of desired signals  $\mathbf{r}_5$  and  $\mathbf{r}_6$ . (b) *Signal Flow Graph*: Linear processing of source signals  $\{\mathbf{x}_1, \mathbf{x}_2\}$  results in signals transmitted along edges of the graph.

Let  $(\cdot)^T$ ,  $(\cdot)^{-1}$ , and  $\text{tr}(\cdot)$  denote matrix transpose, inverse, and trace respectively. Let  $\mathbf{A} \otimes \mathbf{B}$  denote the Kronecker matrix product of two matrices. The matrix  $\mathbf{I}_\ell$  denotes the  $\ell \times \ell$  identity. For  $\ell \geq k$ , the notation  $\mathbf{T}_{k:\ell} \triangleq \mathbf{T}_k \mathbf{T}_{k+1} \cdots \mathbf{T}_\ell$  denotes the product of  $(\ell - k + 1)$  matrices. A matrix  $\mathbf{X} \in \mathbb{R}^{m \times n}$  is written in vector form  $\text{vec}(\mathbf{X}) \in \mathbb{R}^{mn}$  by stacking its columns; i.e.  $\text{vec}(\mathbf{X}) = [\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_n]$  where  $\mathbf{x}_j$  is the  $j$ -th column of  $\mathbf{X}$ . For random vectors,  $E[\cdot]$  denotes the expectation, and  $\Sigma_{\mathbf{x}} \triangleq E[\mathbf{x}\mathbf{x}^T]$  denotes the covariance matrix of the zero-mean random vector  $\mathbf{x}$ .

## II. PROBLEM STATEMENT

Fig. 1 serves as an extended example of an LTN graph. The network is comprised of two sources, two relays, and two receiver nodes.

*Definition 1 (Relay Network)*: Consider a relay network modeled by a directed acyclic graph (DAG)  $G = (\mathcal{V}, \mathcal{E})$  and a set of weights  $\mathcal{C}$ . The set  $\mathcal{V} = \{v_1, v_2, \dots, v_{|\mathcal{V}|}\}$  is the vertex/node set,  $\mathcal{E} \subset \{1, \dots, |\mathcal{V}|\} \times \{1, \dots, |\mathcal{V}|\}$  is the edge set, and  $\mathcal{C} = \{c_{ij} \in \mathbb{Z}^+ : (i, j) \in \mathcal{E}\}$  is the set of weights. Each edge  $(i, j) \in \mathcal{E}$  represents a communication link with integer bandwidth  $c_{ij}$  from node  $v_i$  to  $v_j$ . The in-degree and out-degree

of a node  $v_i$  are computed as

$$d_i^- = \sum_{q:(q,i) \in \mathcal{E}} c_{qi}, \quad (1)$$

$$d_i^+ = \sum_{l:(i,l) \in \mathcal{E}} c_{il}. \quad (2)$$

As an example, the graph in Fig. 1 consists of nodes  $\mathcal{V} = \{v_1, v_2, \dots, v_6\}$ . Integer bandwidths  $c_{ij}$  for each communication link  $(i, j)$  are marked.

*Definition 2 (Source and Receiver Nodes):* Given a relay network  $G = (\mathcal{V}, \mathcal{E})$ , the set of source nodes  $\mathcal{S} \subset \mathcal{V}$  is defined as  $\mathcal{S} = \{v_i \in \mathcal{V} \mid d_i^- = 0\}$ . We assume a labeling of nodes in  $\mathcal{V}$  so that  $\mathcal{S} = \{v_1, v_2, \dots, v_{|\mathcal{S}|}\}$ , i.e. the first  $|\mathcal{S}|$  nodes are source nodes. The set of receiver nodes  $\mathcal{T} \subset \mathcal{V}$  is defined as  $\mathcal{T} = \{v_i \in \mathcal{V} \mid d_i^+ = 0\}$ .<sup>1</sup> Let  $\kappa \triangleq |\mathcal{V}| - |\mathcal{T}|$ . We assume a labeling of nodes in  $\mathcal{V}$  so that  $\mathcal{T} = \{v_{\kappa+1}, v_{\kappa+2}, \dots, v_{|\mathcal{V}|}\}$ , i.e. the last  $|\mathcal{T}|$  nodes are receiver nodes.

In Fig. 1,  $\mathcal{S} = \{v_1, v_2\}$  and  $\mathcal{T} = \{v_5, v_6\}$ .

#### A. Source Model

*Definition 3 (Basic Source Model):* Given a relay network  $G = (\mathcal{V}, \mathcal{E})$  with source/receiver nodes  $(\mathcal{S}, \mathcal{T})$ , the source nodes  $\mathcal{S} = \{v_i\}_{i=1}^{|\mathcal{S}|}$  observe random signals  $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^{|\mathcal{S}|}$ . The random vectors  $\mathbf{x}_i \in \mathbb{R}^{n_i}$  are assumed zero-mean with covariance  $\Sigma_{ii}$ , and cross-covariances  $\Sigma_{ij} \in \mathbb{R}^{n_i \times n_j}$ . Let  $n \triangleq \sum_i n_i$ . The distributed network sources may be grouped into an  $n$ -dimensional random vector  $\mathbf{x} = [\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_{|\mathcal{S}|}]$  with known second-order statistics  $\Sigma_{\mathbf{x}} \in \mathbb{R}^{n \times n}$ ,

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1|\mathcal{S}|} \\ \Sigma_{21} & \Sigma_{22} & \dots & \Sigma_{2|\mathcal{S}|} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{|\mathcal{S}|1} & \Sigma_{|\mathcal{S}|2} & \dots & \Sigma_{|\mathcal{S}||\mathcal{S}|} \end{bmatrix}. \quad (3)$$

More generally, each source node  $v_i \in \mathcal{S}$  emits independent and identically distributed (*i.i.d.*) source vectors  $\{\mathbf{x}_i[t]\}_{t \geq 0}$  for  $t$  a discrete time index; however, in the analysis of zero-delay linear coding, we do not write the time indices explicitly.

*Remark 1:* A common linear signal-plus-noise model for sensor networks is of the form  $\mathbf{x}_i = \mathbf{H}_i \mathbf{x} + \mathbf{n}_i$ ; however, neither a linear source model nor the specific distribution of  $\mathbf{x}_i$  is assumed here. *A priori* knowledge of second-order statistics may be obtained during a training phase via sample estimation.

<sup>1</sup>For networks of interest in this paper, an arbitrary DAG  $G$  may be augmented with auxiliary nodes to ensure that source nodes have in-degree  $d_i^- = 0$  and receiver nodes have out-degree  $d_i^+ = 0$ .

In Fig. 1, two source nodes  $\mathcal{S} = \{v_1, v_2\}$  observe the corresponding random signals in  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2\}$ .

### B. Communication Model

*Definition 4 (Communication Model):* Given a relay network  $G = (\mathcal{V}, \mathcal{E})$  with weight-set  $\mathcal{C}$ , each edge  $(i, j) \in \mathcal{E}$  represents a communication link of bandwidth  $c_{ij}$  from  $v_i$  to  $v_j$ . The bandwidth is the dimension of the vector channel. We denote signals exiting  $v_i \in \mathcal{V}$  along edge  $(i, j) \in \mathcal{E}$  by  $\mathbf{x}_{ij} \in \mathbb{R}^{c_{ij}}$  and signals entering node  $v_j$  along edge  $(i, j) \in \mathcal{E}$  by  $\mathbf{y}_{ij} \in \mathbb{R}^{c_{ij}}$ . If communication is noiseless,  $\mathbf{y}_{ij} = \mathbf{x}_{ij}$ . For all relay nodes and receiver nodes, we further define  $\mathbf{y}_j \in \mathbb{R}^{d_j^-}$  to be the concatenation of all signals  $\mathbf{y}_{ij}$  incident to node  $v_j$  along edges  $(i, j) \in \mathcal{E}$ .

A noisy communication link  $(i, j) \in \mathcal{E}$  is modeled as:  $\mathbf{y}_{ij} = \mathbf{x}_{ij} + \mathbf{z}_{ij}$ . The channel noise  $\mathbf{z}_{ij} \in \mathbb{R}^{c_{ij}}$  is a Gaussian random vector with zero-mean and covariance  $\mathbf{\Sigma}_{\mathbf{z}_{ij}}$ . The channel input is power constrained so that  $E[\|\mathbf{x}_{ij}\|_2^2] \leq P_{ij}$ . The power constraints for a network are given by set  $\mathcal{P} = \{P_{ij} \in \mathbb{R}^+ : (i, j) \in \mathcal{E}\}$ . The signal-to-noise ratio (SNR) along a link is

$$SNR_{ij} = \frac{E[\|\mathbf{x}_{ij}\|_2^2]}{E[\|\mathbf{z}_{ij}\|_2^2]}. \quad (4)$$

Fig. 1(b) illustrates the signal flow of an LTN graph.

### C. Linear Encoding over Graph $G$

Source and relay nodes encode random vector signals by applying reduced-dimension linear transforms.

*Definition 5 (Linear Encoding):* Given a relay network  $G = (\mathcal{V}, \mathcal{E})$ , weight-set  $\mathcal{C}$ , source/receiver nodes  $(\mathcal{S}, \mathcal{T})$ , sources  $\mathcal{X}$ , and the communication model of Definition 4, the linear encoding matrices for  $G$  are denoted by set  $\mathcal{L}_G = \{\mathbf{L}_{ij} : (i, j) \in \mathcal{E}\}$ . Each  $\mathbf{L}_{ij}$  represents the linear transform applied by node  $v_i$  in communication with node  $v_j$ . For  $v_i \in \mathcal{S}$ , transform  $\mathbf{L}_{ij}$  is of size  $c_{ij} \times n_i$  and represents the encoding  $\mathbf{x}_{ij} = \mathbf{L}_{ij}\mathbf{x}_i$ . For a relay  $v_i$ , transform  $\mathbf{L}_{ij}$  is of size  $c_{ij} \times d_i^-$ , and  $\mathbf{x}_{ij} = \mathbf{L}_{ij}\mathbf{y}_i$ . The *compression ratio* along edge  $(i, j) \in \mathcal{E}$  is

$$\alpha_{ij} = \begin{cases} \frac{c_{ij}}{n_i} & \text{if } v_i \in \mathcal{S}, \\ \frac{c_{ij}}{d_i^-} & \text{if } v_i \in \mathcal{V} \setminus \mathcal{S}. \end{cases} \quad (5a)$$

$$(5b)$$

In Fig. 1, the linear encoding matrices for source node  $v_1$  and  $v_2$  are  $\{\mathbf{L}_{15}, \mathbf{L}_{13}\}$  and  $\{\mathbf{L}_{26}, \mathbf{L}_{23}\}$  respectively. The linear encoding matrices for the relays are  $\mathbf{L}_{34}$ ,  $\mathbf{L}_{45}$ ,  $\mathbf{L}_{46}$ . The output signals of source node  $v_1$  are

$\mathbf{x}_{15} = \mathbf{L}_{15}\mathbf{x}_1$  and  $\mathbf{x}_{13} = \mathbf{L}_{13}\mathbf{x}_1$ . Similarly, the output signal of relay  $v_3$  is

$$\mathbf{x}_{34} = \mathbf{L}_{34}\mathbf{y}_3 = \mathbf{L}_{34} \begin{bmatrix} \mathbf{y}_{13} \\ \mathbf{y}_{23} \end{bmatrix}. \quad (6)$$

#### D. Linear Estimation over $G$

*Definition 6 (Linear Estimation):* Given relay network  $G = (\mathcal{V}, \mathcal{E})$ , weight-set  $\mathcal{C}$ , source/receiver nodes  $(\mathcal{S}, \mathcal{T})$ , sources  $\mathcal{X}$ , and the communication model of Def. 4, the set of linear decoding matrices is denoted  $\mathcal{B}_G = \{\mathbf{B}_i\}_{i:v_i \in \mathcal{T}}$ . Each receiver  $v_i \in \mathcal{T}$  estimates a (zero-mean) random vector  $\mathbf{r}_i \in \mathbb{R}^{r_i}$  which is correlated with the sources in  $\mathcal{X}$ . We assume that the second-order statistics  $\Sigma_{\mathbf{r}_i}$ ,  $\Sigma_{\mathbf{r}_i\mathbf{x}}$  are known. Receiver  $v_i \in \mathcal{T}$  applies a linear estimator given by matrix  $\mathbf{B}_i \in \mathbb{R}^{r_i \times d_i^-}$  to estimate  $\mathbf{r}_i$  given its observations and computes  $\hat{\mathbf{r}}_i = \mathbf{B}_i\mathbf{y}_i$ . The linear least squares estimate (LLSE) of  $\mathbf{r}_i$  is denoted by  $\hat{\mathbf{r}}_i$ .

In Fig. 1, receiver  $v_5$  reconstructs  $\mathbf{r}_5$  while receiver  $v_6$  reconstructs  $\mathbf{r}_6$ . The LLSE signals  $\hat{\mathbf{r}}_5$  and  $\hat{\mathbf{r}}_6$  are computed as

$$\hat{\mathbf{r}}_5 = \mathbf{B}_5\mathbf{y}_5 = \mathbf{B}_5 \begin{bmatrix} \mathbf{y}_{15} \\ \mathbf{y}_{45} \end{bmatrix}, \quad (7)$$

$$\hat{\mathbf{r}}_6 = \mathbf{B}_6\mathbf{y}_6 = \mathbf{B}_6 \begin{bmatrix} \mathbf{y}_{26} \\ \mathbf{y}_{46} \end{bmatrix}. \quad (8)$$

*Definition 7 (Distortion Metric):* Let  $\mathbf{x}$  and  $\mathbf{y}$  be two real vectors of the same dimension. The MSE distortion metric is defined as

$$d_{mse}(\mathbf{x}, \mathbf{y}) \triangleq \|\mathbf{x} - \mathbf{y}\|_2^2. \quad (9)$$

#### E. Compression-Estimation in Networks

*Definition 8 (Linear Transform Network  $\mathcal{N}$ ):* An LTN model  $\mathcal{N}$  is a communication network modeled by DAG  $G = (\mathcal{V}, \mathcal{E})$ , weight-set  $\mathcal{C}$ , source/receiver nodes  $(\mathcal{S}, \mathcal{T})$ , sources  $\mathcal{X}$ , sets  $\mathcal{L}_G$ , and  $\mathcal{B}_G$  from Definitions 1-6. Second-order source statistics are given by  $\Sigma_{\mathbf{x}}$  (Definition 3). The operational meaning of compression-estimation matrices in  $\mathcal{L}_G$  and  $\mathcal{B}_G$  is in terms of signal flows on  $G$  (Definition 4). The desired reconstruction vectors  $\{\mathbf{r}_i\}_{i:v_i \in \mathcal{T}}$  have known second-order statistics  $\Sigma_{\mathbf{r}_i}$  and  $\Sigma_{\mathbf{r}_i\mathbf{x}}$ . The set  $\{\hat{\mathbf{r}}_i\}_{i:v_i \in \mathcal{T}}$  denotes the LLSE estimates formed at receivers (Definition 6). For noisy networks, noise variables along link  $(i, j) \in \mathcal{E}$  have known covariances  $\Sigma_{\mathbf{z}_{ij}}$ . Power constraints are given by set  $\mathcal{P}$  in Definition 4.

Given an LTN graph  $\mathcal{N}$ , the task is to design a *network transform code*: the compression-estimation matrices in  $\mathcal{L}_G$  and  $\mathcal{B}_G$  to minimize the end-to-end weighted MSE distortion. Let positive weights  $\{w_i\}_{i:v_i \in \mathcal{T}}$  represent



the relative importance of reconstructing a signal at receiver  $v_i \in \mathcal{T}$ . Using indexing term  $\kappa \triangleq |\mathcal{V}| - |\mathcal{T}|$  for receiver nodes, we concatenate vectors  $\mathbf{r}_i$  as  $\mathbf{r} = [\mathbf{r}_{\kappa+1}; \mathbf{r}_{\kappa+2}; \dots; \mathbf{r}_{|\mathcal{V}|}]$  and LLSE estimates  $\hat{\mathbf{r}}_i$  as  $\hat{\mathbf{r}} = [\hat{\mathbf{r}}_{\kappa+1}; \hat{\mathbf{r}}_{\kappa+2}; \dots; \hat{\mathbf{r}}_{|\mathcal{V}|}]$ . The average weighted MSE written via a weighted  $\ell^2$ -norm is

$$\begin{aligned} D_{MSE, \mathbf{W}} &\triangleq E \left[ \sum_{i: v_i \in \mathcal{T}} d_{mse}(\sqrt{w_i} \mathbf{r}_i, \sqrt{w_i} \hat{\mathbf{r}}_i) \right], \\ &= E \left[ \|\mathbf{r} - \hat{\mathbf{r}}\|_{\mathbf{W}}^2 \right], \end{aligned} \quad (10)$$

where  $\mathbf{W}$  contains diagonal blocks  $\mathbf{W}_i = \sqrt{w_i} \mathbf{I}$ .

*Remark 2:* The distortion  $D_{MSE, \mathbf{W}}$  is a function of the compression matrices in  $\mathcal{L}_G$  and the estimation matrices in  $\mathcal{B}_G$ . In most network topologies, the weighted MSE distortion is non-convex over the set of feasible matrices. Even in the particular case of distributed compression [5], currently the optimal linear transforms are not solvable in closed form.

### III. LINEAR SIGNAL PROCESSING IN NETWORKS

The linear processing and filtering of source signals by an LTN graph  $\mathcal{N}$  is modeled compactly as a linear system with inputs, outputs, and memory elements. At each time step, LTN nodes transmit random signals through edges/channels of the graph.

#### A. Linear System

Consider edge  $(i, j) \in \mathcal{E}$  as a memory element storing random vector  $\mathbf{y}_{ij}$ . Let  $c \triangleq (\sum_{(i,j) \in \mathcal{E}} c_{ij})$  and  $d \triangleq (\sum_{i: v_i \in \mathcal{T}} d_i^-)$ . The network  $\mathcal{N}$  is modeled as a linear system with the following signals: (i) input sources  $\{\mathbf{x}_i\}_{i: v_i \in \mathcal{S}}$  concatenated as global source vector  $\mathbf{x} \in \mathbb{R}^n$ ; (ii) input noise variables  $\{\mathbf{z}_{ij}\}_{(i,j) \in \mathcal{E}}$  concatenated as global noise vector  $\mathbf{z} \in \mathbb{R}^c$ ; (iii) memory elements  $\{\mathbf{y}_{ij}\}_{(i,j) \in \mathcal{E}}$  concatenated as global state vector  $\boldsymbol{\mu}[t] \in \mathbb{R}^c$  at time  $t$ ; (iv) output vectors  $\{\mathbf{y}_i\}_{i: v_i \in \mathcal{T}}$  concatenated as  $\mathbf{y} \in \mathbb{R}^d$ .

1) *State-space Equations:* The linear system<sup>2</sup> is described by the following state-space equations for  $i: v_i \in \mathcal{T}$ ,

$$\boldsymbol{\mu}[t+1] = \mathbf{F}\boldsymbol{\mu}[t] + \mathbf{E}\mathbf{x}[t] + \tilde{\mathbf{E}}\mathbf{z}[t], \quad (11)$$

$$\mathbf{y}_i[t] = \mathbf{C}_i\boldsymbol{\mu}[t] + \mathbf{D}_i\mathbf{x}[t] + \tilde{\mathbf{D}}_i\mathbf{z}[t]. \quad (12)$$

The matrix  $\mathbf{F} \in \mathbb{R}^{c \times c}$  is the state-evolution matrix common to all receivers,  $\mathbf{E} \in \mathbb{R}^{c \times n}$  is the source-network connectivity matrix, and  $\tilde{\mathbf{E}} \in \mathbb{R}^{c \times c}$  is the noise-to-network connectivity matrix. The matrices  $\mathbf{C}_i \in \mathbb{R}^{d_i^- \times c}$ ,

<sup>2</sup>When discussing zero-delay linear coding, the time indices on vectors  $\mathbf{x}$ ,  $\mathbf{z}$ , and  $\mathbf{y}_i$  are omitted for greater clarity of presentation.

$\mathbf{D}_i \in \mathbb{R}^{d_i^- \times n}$ , and  $\tilde{\mathbf{D}}_i \in \mathbb{R}^{d_i^- \times c}$  represent how each receiver's output is related to the state, source, and noise vectors respectively. For networks considered in this paper,  $\mathbf{D}_i = \mathbf{0}$  and  $\tilde{\mathbf{D}}_i = \mathbf{0}$ .

2) *Linear Transfer Function*: A standard result in linear system theory yields the transfer function (assuming a unity indeterminate delay operator) for each receiver  $v_i \in \mathcal{T}$ ,

$$\mathbf{y}_i = \mathbf{C}_i (\mathbf{I} - \mathbf{F})^{-1} (\mathbf{E}\mathbf{x} + \tilde{\mathbf{E}}\mathbf{z}), \quad (13)$$

$$= \mathbf{G}_i \mathbf{x} + \tilde{\mathbf{G}}_i \mathbf{z}, \quad (14)$$

where  $\mathbf{G}_i \triangleq \mathbf{C}_i (\mathbf{I} - \mathbf{F})^{-1} \mathbf{E}$  and  $\tilde{\mathbf{G}}_i \triangleq \mathbf{C}_i (\mathbf{I} - \mathbf{F})^{-1} \tilde{\mathbf{E}}$ . For acyclic graphs,  $\mathbf{F}$  is a nilpotent matrix and  $(\mathbf{I} - \mathbf{F})^{-1} = \mathbf{I} + \sum_{k=1}^{\gamma} \mathbf{F}^k$  for finite integer  $\gamma$ . Using indexing term  $\kappa$ , the observation vectors collected by receivers are concatenated as  $\mathbf{y} = [\mathbf{y}_{\kappa+1}; \mathbf{y}_{\kappa+2}; \dots; \mathbf{y}_{|\mathcal{V}|}]$ . Let

$$\mathbf{T} \triangleq [\mathbf{G}_{\kappa+1}; \mathbf{G}_{\kappa+2}; \dots; \mathbf{G}_{|\mathcal{V}|}], \quad (15)$$

and let  $\tilde{\mathbf{T}}$  be defined similarly with respect to matrices  $\tilde{\mathbf{G}}_i$ . Then the complete linear transfer function of the network  $\mathcal{N}$  is  $\mathbf{y} = \mathbf{T}\mathbf{x} + \tilde{\mathbf{T}}\mathbf{z}$ . Analog processing of signals without error control implies noise propagation; the additive noise  $\mathbf{z}$  is also linearly filtered by the network via  $\tilde{\mathbf{T}}$ .

*Example 1*: Fig. 2 is the LTN graph of a noisy relay network. Let state  $\boldsymbol{\mu} = [\mathbf{y}_{12}; \mathbf{y}_{13}; \mathbf{y}_{23}]$ ,  $\mathbf{z} = [\mathbf{z}_{12}; \mathbf{z}_{13}; \mathbf{z}_{23}]$ , and output  $\mathbf{y}_3 = [\mathbf{y}_{13}; \mathbf{y}_{23}]$ . The linear system representation is given as follows,

$$\begin{aligned} \boldsymbol{\mu}[t+1] &= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{L}_{23} & \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{\mu}[t] + \begin{bmatrix} \mathbf{L}_{12} \\ \mathbf{L}_{13} \\ \mathbf{0} \end{bmatrix} \mathbf{x}_1[t] + \mathbf{I}_c \mathbf{z}[t], \\ \mathbf{y}_3[t] &= \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \boldsymbol{\mu}[t]. \end{aligned}$$

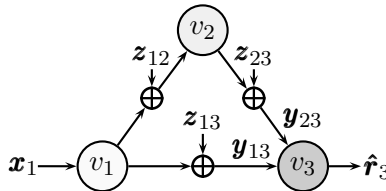
By evaluating Eqn. (14),

$$\mathbf{y}_3[t] = \begin{bmatrix} \mathbf{L}_{13} \\ \mathbf{L}_{23}\mathbf{L}_{12} \end{bmatrix} \mathbf{x}_1[t] + \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{L}_{23} & \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{z}[t].$$

Dropping the time indices and writing  $\mathbf{x} = \mathbf{x}_1$  in addition to  $\mathbf{y} = \mathbf{y}_3$ , the linear transfer function of the noisy relay network is of the following form:  $\mathbf{y} = \mathbf{T}\mathbf{x} + \tilde{\mathbf{T}}\mathbf{z}$ .

### B. Layered Networks

*Definition 9 (Layered DAG Network)*: A layering of a DAG  $G = (\mathcal{V}, \mathcal{E})$  is a partition of  $\mathcal{V}$  into disjoint subsets  $\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_{p+1}$  such that if directed edge  $(u, v) \in \mathcal{E}$ , where  $u \in \mathcal{V}_j$  and  $v \in \mathcal{V}_k$ , then  $j > k$ . A DAG layering (non-unique) is polynomial-time computable [32].



Given a layered partition  $\{\mathcal{V}_\ell\}_{\ell=1}^{p+1}$  of an LTN graph, source nodes  $v_i \in \mathcal{S}$  with in-degree  $d_i^- = 0$  may be placed in partition  $\mathcal{V}_{p+1}$ . Similarly, receivers  $v_i \in \mathcal{T}$  with out-degree  $d_i^+ = 0$  may be placed in partition  $\mathcal{V}_1$ . The transfer function  $\mathbf{T}$  in Eqn. (15) may be factored into a product of matrices,

$$\mathbf{T} = \mathbf{T}_{1:p} \triangleq \mathbf{T}_1 \mathbf{T}_2 \cdots \mathbf{T}_p, \quad (16)$$

where  $\mathbf{T}_\ell$  for  $1 \leq \ell \leq p$  is the linear transformation of signals between nodes in partition  $\mathcal{V}_{\ell+1}$  and  $\mathcal{V}_\ell$  (note the reverse ordering of the  $\mathbf{T}_\ell$  with respect to the partitions  $\mathcal{V}_\ell$ ). If an edge exists between nodes in non-consecutive partitions, an identity transform is inserted to replicate signals between multiple layers. Due to the linearity of transforms, for any layered partition  $\{\mathcal{V}_\ell\}_{\ell=1}^{p+1}$  of  $\mathcal{V}$ , the layered transforms  $\{\mathbf{T}_\ell\}_{\ell=1}^p$  can be constructed. The  $\{\mathbf{T}_\ell\}_{\ell=1}^p$  are structured matrices comprised of sub-blocks  $\mathbf{L}_{ij}$ , identity matrices, and/or zero matrices. The block structure is determined by the network topology.

*Example 2:* For the multiple unicast network of Fig. 1, a valid layered partition of  $\mathcal{V}$  is  $\mathcal{V}_1 = \{v_5, v_6\}$ ,  $\mathcal{V}_2 = \{v_4\}$ ,  $\mathcal{V}_3 = \{v_3\}$ , and  $\mathcal{V}_4 = \{v_1, v_2\}$ . Let  $\mathbf{x} = [\mathbf{x}_1; \mathbf{x}_2]$ ,  $\mathbf{y} = [\mathbf{y}_5; \mathbf{y}_6] = [\mathbf{y}_{15}; \mathbf{y}_{45}; \mathbf{y}_{46}; \mathbf{y}_{26}]$ , and let  $\mathbf{L}_{34}$  be partitioned as  $\mathbf{L}_{34} = [\mathbf{L}'_{34} \quad \mathbf{L}''_{34}]$ . According to the layering, the transfer matrix  $\mathbf{T}$  is factored in product form  $\mathbf{T} = \mathbf{T}_1 \mathbf{T}_2 \mathbf{T}_3$ .

$$\mathbf{T} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{45} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{46} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}'_{34} & \mathbf{L}''_{34} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{L}_{15} & \mathbf{0} \\ \mathbf{L}_{13} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{23} \\ \mathbf{0} & \mathbf{L}_{26} \end{bmatrix}.$$

*Example 3:* Consider the setting of Example 1 for the relay network shown in Fig. 2. A valid layered partition of  $\mathcal{V}$  is  $\mathcal{V}_1 = \{v_3\}$ ,  $\mathcal{V}_2 = \{v_2\}$ ,  $\mathcal{V}_3 = \{v_1\}$ . According to the layering, the transfer matrix  $\mathbf{T}$  may

be written in product form  $\mathbf{T} = \mathbf{T}_1 \mathbf{T}_2$ ,

$$\mathbf{T} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{23} \end{bmatrix} \begin{bmatrix} \mathbf{L}_{13} \\ \mathbf{L}_{12} \end{bmatrix}.$$

#### IV. OPTIMIZING COMPRESSION-ESTIMATION MATRICES

Our optimization method proceeds iteratively over network layers. To simplify the optimization, we first assume ideal channels (high-SNR communication) for which  $\mathbf{y}_{ij} = \mathbf{x}_{ij}$ . Then the linear operation of the network  $\mathcal{N}$  is  $\mathbf{y} = \mathbf{T}\mathbf{x}$  with  $\mathbf{z} = \mathbf{0}$ . Linear transform coding is constrained according to bandwidth compression ratios  $\alpha_{ij}$ .

##### A. MSE Distortion at Receivers

According to the linear system equations, Eqns. (11)-(14), each receiver  $v_i \in \mathcal{T}$  receives filtered source observations  $\mathbf{y}_i = \mathbf{G}_i \mathbf{x}$ . Receiver  $v_i$  applies a linear estimator  $\mathbf{B}_i$  to estimate signal  $\mathbf{r}_i$ . The MSE cost of estimation is

$$\begin{aligned} D_i &= E \left[ \|\mathbf{r}_i - \mathbf{B}_i \mathbf{G}_i \mathbf{x}\|_2^2 \right] \\ &= \text{tr}(\Sigma_{\mathbf{r}_i}) - 2\text{tr}(\mathbf{B}_i \mathbf{G}_i \Sigma_{\mathbf{x} \mathbf{r}_i}) + \text{tr}(\mathbf{B}_i \mathbf{G}_i \Sigma_{\mathbf{x}} \mathbf{G}_i^T \mathbf{B}_i^T). \end{aligned} \quad (17)$$

Setting the matrix derivative with respect to  $\mathbf{B}_i$  in Eqn. (17) to zero yields:  $-2\Sigma_{\mathbf{r}_i \mathbf{x}} \mathbf{G}_i^T + 2\mathbf{B}_i \mathbf{G}_i \Sigma_{\mathbf{x}} \mathbf{G}_i^T = 0$ . For a fixed transfer function  $\mathbf{G}_i$ , the optimal LLSE matrix  $\mathbf{B}_i^{\text{opt}}$  is

$$\mathbf{B}_i^{\text{opt}} = \Sigma_{\mathbf{r}_i \mathbf{x}} \mathbf{G}_i^T [\mathbf{G}_i \Sigma_{\mathbf{x}} \mathbf{G}_i^T]^{-1}. \quad (18)$$

If  $\mathbf{G}_i$  in Eqn. (18) is singular, the inverse may be replaced with a pseudo-inverse operation to compute  $\mathbf{B}_i^{\text{opt}}$ .

Let  $\mathbf{B}$  denote a block diagonal global matrix containing individual decoding matrices  $\{\mathbf{B}_i\}_{i:v_i \in \mathcal{T}}$  on the diagonal. For an LTN graph  $\mathcal{N}$  with encoding transfer function  $\mathbf{T} = \mathbf{T}_{1:p}$ , we write the linear decoding operation of all receivers as  $\hat{\mathbf{r}} = \mathbf{B}\mathbf{y}$  where  $\mathbf{y} = \mathbf{T}_{1:p}\mathbf{x}$  are the observations received. The weighted MSE cost in Eqn. (10) for reconstructing signals  $\{\mathbf{r}_i\}_{i:v_i \in \mathcal{T}}$  at all receivers is written as

$$\begin{aligned} D_{MSE, \mathbf{W}} &= E \left[ \|\mathbf{r} - \hat{\mathbf{r}}\|_{\mathbf{W}}^2 \right] \\ &= E \left[ \|\mathbf{r} - \mathbf{B} \mathbf{T}_{1:p} \mathbf{x}\|_{\mathbf{W}}^2 \right] \\ &= \text{tr}(\mathbf{W} \Sigma_{\mathbf{r}} \mathbf{W}^T) - 2\text{tr}(\mathbf{W} \mathbf{B} \mathbf{T}_{1:p} \Sigma_{\mathbf{x} \mathbf{r}} \mathbf{W}^T) \\ &\quad + \text{tr}(\mathbf{W} \mathbf{B} \mathbf{T}_{1:p} \Sigma_{\mathbf{x}} \mathbf{T}_{1:p}^T \mathbf{B}^T \mathbf{W}^T). \end{aligned} \quad (19)$$

By construction of the weighting matrix  $\mathbf{W}$ , the MSE in Eqn. (19) is a weighted sum of individual distortions at receivers, i.e.  $D_{MSE, \mathbf{W}} = \sum_{i:v_i \in \mathcal{T}} w_i D_i$ .

### B. Computing Encoding Transforms $\mathbf{T}_i$

The optimization of the network transfer function  $\mathbf{T} = \mathbf{T}_{1:p}$  is more complex due to block constraints imposed by the network topology on matrices  $\{\mathbf{T}_i\}_{i=1}^p$ . In order to solve for a particular linear transform  $\mathbf{T}_i$ , we assume all linear transforms  $\mathbf{T}_j$ ,  $j \neq i$  and the receivers' decoding transform  $\mathbf{B}$  are fixed. Then the optimal  $\mathbf{T}_i$  is the solution to a constrained quadratic program. To derive this, we utilize the following identities in which  $\mathbf{x} = \text{vec}(\mathbf{X})$ :

$$\text{tr}(\mathbf{A}^T \mathbf{X}) = \text{vec}(\mathbf{A})^T \mathbf{x}, \quad (20)$$

$$\text{tr}(\mathbf{X}^T \mathbf{A}_1 \mathbf{X} \mathbf{A}_2) = \mathbf{x}^T (\mathbf{A}_2 \otimes \mathbf{A}_1) \mathbf{x}. \quad (21)$$

We write the network's linear transfer function as  $\mathbf{T} = \mathbf{T}_{1:p} = \mathbf{T}_{1:i-1} \mathbf{T}_i \mathbf{T}_{i+1:p}$  and define the following matrices

$$\mathbf{J}_i \triangleq \mathbf{T}_{i+1:p} \Sigma_{\mathbf{r}} \mathbf{W}^T \mathbf{W} \mathbf{B} \mathbf{T}_{1:i-1}, \quad (22)$$

$$\mathbf{J}'_i \triangleq (\mathbf{T}_{1:i-1})^T \mathbf{B}^T \mathbf{W}^T \mathbf{W} \mathbf{B} \mathbf{T}_{1:i-1}, \quad (23)$$

$$\mathbf{J}''_i \triangleq \mathbf{T}_{i+1:p} \Sigma_{\mathbf{x}} (\mathbf{T}_{i+1:p})^T. \quad (24)$$

To write  $D_{MSE, \mathbf{W}}$  in terms of the matrix variable  $\mathbf{T}_i$ , we also define the following,

$$p_i \triangleq \text{tr}(\mathbf{W} \Sigma_{\mathbf{r}} \mathbf{W}^T), \quad (25)$$

$$\mathbf{p}_i \triangleq -2 \text{vec}(\mathbf{J}_i^T), \quad (26)$$

$$\mathbf{P}_i \triangleq \mathbf{J}''_i \otimes \mathbf{J}'_i, \quad (27)$$

where  $p_i$ ,  $\mathbf{p}_i$ , and  $\mathbf{P}_i$  are a scalar, vector, and positive semi-definite matrix respectively. The following lemma expresses  $D_{MSE, \mathbf{W}}$  as a function of the unknown matrix variable  $\mathbf{T}_i$ .

*Lemma 1:* Let transforms  $\mathbf{T}_j$ ,  $j \neq i$ , and  $\mathbf{B}$  be fixed. Let  $\mathbf{J}_i$ ,  $\mathbf{J}'_i$ ,  $\mathbf{J}''_i$  be defined in Eqns. (22)-(24), and  $p_i$ ,  $\mathbf{p}_i$ , and  $\mathbf{P}_i$  be defined in Eqns. (25)-(27). Then the weighted MSE distortion  $D_{MSE, \mathbf{W}}$  of Eqn. (19) is a quadratic function of  $\mathbf{t}_i = \text{vec}(\mathbf{T}_i)$ ,

$$D_{MSE, \mathbf{W}} = \mathbf{t}_i^T \mathbf{P}_i \mathbf{t}_i + \mathbf{p}_i^T \mathbf{t}_i + p_i. \quad (28)$$

*Proof:* Substituting the expressions for  $\mathbf{J}_i$ ,  $\mathbf{J}'_i$ ,  $\mathbf{J}''_i$  in Eqns. (22)-(24) into Eqn. (19) produces the intermediate equation:  $D_{MSE, \mathbf{W}} = \text{tr}(\mathbf{T}_i^T \mathbf{J}'_i \mathbf{T}_i \mathbf{J}''_i) - 2 \text{tr}(\mathbf{J}_i \mathbf{T}_i) + p_i$ . Directly applying the vector-matrix identities of Eqns. (20)-(21) results in Eqn. (28). ■

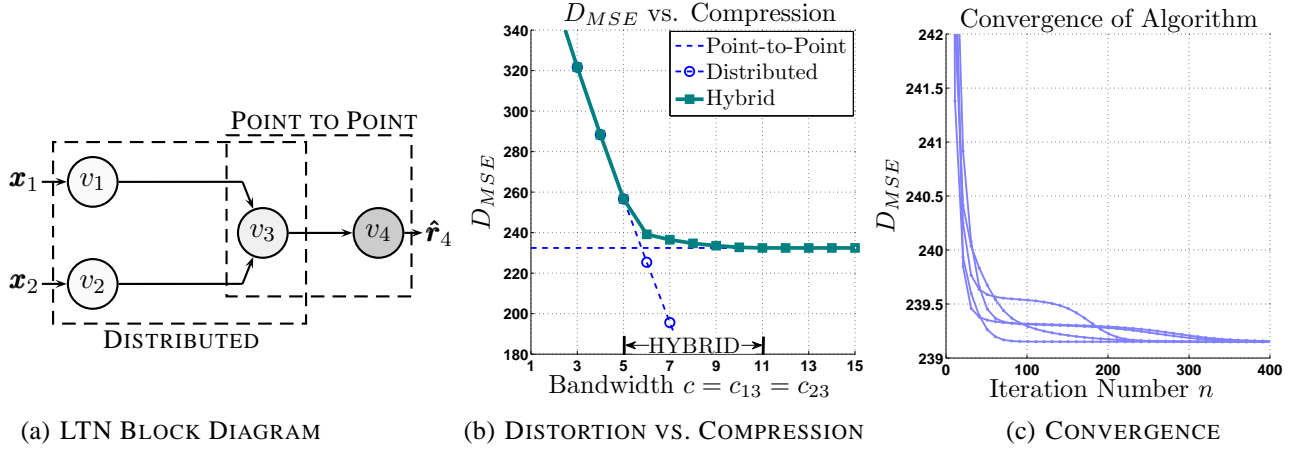


Fig. 3. (a) Block diagram of the “hybrid network” example. (b) The end-to-end distortion vs. compression for varying bandwidth  $c = c_{13} = c_{23}$ . The network operates in one of three modes (distributed, hybrid, or point-to-point) as described in Example 4. (c) Convergence of  $D_{MSE}(n)$  for five different initializations of the iterative algorithm for the operating point  $c = 6, c_{34} = 11$ .

### C. Quadratic Program with Convex Constraints

Due to Lemma 1, the weighted MSE is a quadratic function of  $\mathbf{t}_i = \text{vec}(\mathbf{T}_i)$  if all other network matrices are fixed. The optimal  $\mathbf{T}_i$  must satisfy block constraints determined by network topology. The block constraints are *linear equality constraints* of the form  $\Phi_i \mathbf{t}_i = \phi_i$ . For example, if  $\mathbf{T}_i$  contains an identity sub-block, this is enforced by setting entries in  $\mathbf{t}_i$  to zero and one accordingly, via linear equality constraints.

*Theorem 1 (Optimal Encoding):* Let encoding matrices  $\mathbf{T}_j$ ,  $j \neq i$  and decoding matrix  $\mathbf{B}$  be fixed. Let  $\mathbf{t}_i = \text{vec}(\mathbf{T}_i)$ . The optimal encoding transform  $\mathbf{t}_i$  is given by the following constrained quadratic program (QP) [33, Def. 4.34]

$$\begin{aligned} \arg \min_{\mathbf{t}_i} \quad & \mathbf{t}_i^T \mathbf{P}_i \mathbf{t}_i + \mathbf{p}_i^T \mathbf{t}_i + p_i \\ \text{s. t.} \quad & \Phi_i \mathbf{t}_i = \phi_i, \end{aligned} \quad (29)$$

where  $(\Phi_i, \phi_i)$  represent linear equality constraints on elements of  $\mathbf{T}_i$ . The solution to the above optimization for  $\mathbf{t}_i$  is obtained by solving a corresponding linear system

$$\begin{bmatrix} 2\mathbf{P}_i & \Phi_i^T \\ \Phi_i & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{t}_i \\ \lambda \end{bmatrix} = \begin{bmatrix} -\mathbf{p}_i \\ \phi_i \end{bmatrix}. \quad (30)$$

If the constraints determined by the pair  $(\Phi_i, \phi_i)$  are feasible, the linear system of Eqn. (30) is guaranteed to have either one or infinitely many solutions.

**Algorithm 1** IDEAL-COMPRESSION-ESTIMATION( $\mathcal{N}$ ,  $\mathbf{W}$ ,  $\epsilon$ )

- 
- 1: Identify compression matrices  $\{\mathbf{T}_i\}_{i=1}^p$  and corresponding linear equalities  $\{\Phi_i, \phi_i\}_{i=1}^p$  for network  $\mathcal{N}$ .  
Identify estimation matrices  $\{\mathbf{B}_i\}_{i:v_i \in \mathcal{T}}$ . [Sec. III, Sec. IV-C]
  - 2: Initialize  $\{\mathbf{T}_i^{(0)}\}_{i=1}^p$  randomly to feasible matrices.
  - 3: Set  $n = 1$ ,  $D_{MSE, \mathbf{W}}(0) = \infty$ .
  - 4: **repeat**
  - 5:   Compute  $\{\mathbf{B}_i^{(n)}\}_{i:v_i \in \mathcal{T}}$  given  $\{\mathbf{T}_k^{(n-1)}\}_{k=1}^p$ . [Eqn. (18)]
  - 6:   **for**  $i = 1 : p$  **do**
  - 7:     Compute  $\mathbf{T}_i^{(n)}$  given  $\{\Phi_i, \phi_i\}$ ,  $\{\mathbf{B}_k^{(n)}\}_{k:v_k \in \mathcal{T}}$ ,  $\{\mathbf{T}_k^{(n)}\}_{k=1}^{(i-1)}$ ,  $\{\mathbf{T}_k^{(n-1)}\}_{k=i+1}^p$ . [Theorem 1]
  - 8:   **end for**
  - 9:   Compute  $D_{MSE, \mathbf{W}}(n)$ . [Eqn. (19)]
  - 10:   Set  $\Delta_{MSE, \mathbf{W}} = D_{MSE, \mathbf{W}}(n) - D_{MSE, \mathbf{W}}(n-1)$ .
  - 11:   Set  $n = n + 1$ .
  - 12: **until**  $\Delta_{MSE, \mathbf{W}} \leq \epsilon$  or  $n \geq N_{max}$ .
  - 13: **return**  $\{\mathbf{T}_i^{(n)}\}_{i=1}^p$ ,  $\{\mathbf{B}_i^{(n)}\}_{i:v_i \in \mathcal{T}}$ .
- 

*Proof:* The QP of Eqn. (29) follows from Lemma 1 with additional linear equality constraints placed on  $\mathbf{t}_i$ . The closed form solution to the QP is derived using Lagrange dual multipliers for the linear constraints, and the Karush-Kuhn-Tucker (KKT) conditions. Let  $f(\mathbf{t}_i, \boldsymbol{\lambda})$  represent the Lagrangian formed with dual vector variable  $\boldsymbol{\lambda}$  for the constraints,

$$f(\mathbf{t}_i, \boldsymbol{\lambda}) = \mathbf{t}_i^T \mathbf{P}_i \mathbf{t}_i + \mathbf{p}_i^T \mathbf{t}_i + p_i + \boldsymbol{\lambda}^T (\Phi_i \mathbf{t}_i - \phi_i), \quad (31)$$

$$\nabla_{\mathbf{t}_i} f(\mathbf{t}_i, \boldsymbol{\lambda}) = 2\mathbf{P}_i \mathbf{t}_i + \mathbf{p}_i + \Phi_i^T \boldsymbol{\lambda}, \quad (32)$$

$$\nabla_{\boldsymbol{\lambda}} f(\mathbf{t}_i, \boldsymbol{\lambda}) = \Phi_i \mathbf{t}_i - \phi_i. \quad (33)$$

Setting  $\nabla_{\mathbf{t}_i} f(\mathbf{t}_i, \boldsymbol{\lambda}) = \mathbf{0}$  and  $\nabla_{\boldsymbol{\lambda}} f(\mathbf{t}_i, \boldsymbol{\lambda}) = \mathbf{0}$  yields the linear system of Eqn. (30), the solutions to which are  $\mathbf{t}_i$  and dual vector  $\boldsymbol{\lambda}$ . Since the MSE distortion is bounded by a minimum of zero error, the linear system has a unique solution if  $\mathbf{P}_i$  is full rank, or infinitely many solutions of equivalent objective value if  $\mathbf{P}_i$  is singular. ■

*Remark 3:* Beyond linear constraints, several other convex constraints on matrix variables could be applied within the quadratic program. For example, the  $\ell_1$ -norm of a vector  $\mathbf{x} \in \mathbb{R}^n$  defined by  $\|\mathbf{x}\|_1 \triangleq \sum_i |x_i|$  is often used in compressed sensing to enforce *sparsity*.

TABLE I  
A “HYBRID” LINEAR TRANSFORM NETWORK

Network Modes	Bandwidth
<i>Distributed</i>	$c \leq \lfloor \frac{c_{34}}{2} \rfloor$
<i>Hybrid</i>	$\lceil \frac{c_{34}}{2} \rceil < c < c_{34}$
<i>Point to Point</i>	$c_{34} \leq c$

#### D. An Iterative Algorithm

Algorithm 1 defines an iterative method to optimize all encoding matrices  $\{\mathbf{T}_i\}_{i=1}^p$  and the global decoding matrix  $\mathbf{B}$  for an LTN graph. The iterative algorithm begins with the random initialization of the encoding matrices  $\{\mathbf{T}_i\}_{i=1}^p$  subject to size specifications and linear equality constraints given by  $\{\Phi_i\}_{i=1}^p$  and  $\{\phi_i\}_{i=1}^p$ . The iterative method proceeds by solving for the optimal  $\mathbf{B}$  transform first. Similarly, with  $\mathbf{T}_j, j \neq i$  and  $\mathbf{B}$  fixed, the optimal  $\mathbf{T}_i$  is computed using Theorem 1. The iterative method proceeds for  $n \leq N_{max}$  iterations or until the difference in error  $\Delta_{MSE, \mathbf{W}}$  is less than a prescribed tolerance  $\epsilon$ .

#### E. Convergence to Stationary Points

A key property of Algorithm 1 is the convergence to a stationary point (either local minimum or saddle-point) of the weighted MSE.

*Theorem 2 (Local Convergence):* Denote the network’s linear transfer function after the  $n$ -th outer-loop iteration in Algorithm 1 by  $\mathbf{T}^{(n)}$ , and the block-diagonal global decoding transform by  $\mathbf{B}^{(n)}$  which contains matrices  $\{\mathbf{B}_i^{(n)}\}_{i:v_i \in \mathcal{T}}$  on the diagonal. Let  $\hat{\mathbf{r}}^{(n)} = \mathbf{B}^{(n)}\mathbf{T}^{(n)}\mathbf{x}$  denote the estimate of desired signal  $\mathbf{r}$ . Then

$$E \left[ \left\| \mathbf{r} - \hat{\mathbf{r}}^{(n)} \right\|_{\mathbf{W}}^2 \right] \geq E \left[ \left\| \mathbf{r} - \hat{\mathbf{r}}^{(n+1)} \right\|_{\mathbf{W}}^2 \right], \quad (34)$$

i.e., the weighted MSE distortion is a nonincreasing function of the iteration number  $n$ .

*Proof:* In Step 5 of Algorithm 1, with matrices  $\{\mathbf{T}_k^{(n-1)}\}_{k=1}^p$  fixed, the optimal transform  $\mathbf{B}^{(n)}$  is determined to minimize  $D_{MSE, \mathbf{W}}$ . The current transform  $\mathbf{B}^{(n-1)}$  is feasible within the optimization space which implies that the MSE distortion cannot increase. In Step 7 of the inner loop, with matrices  $\mathbf{B}^{(n)}$ ,  $\{\mathbf{T}_k^{(n)}\}_{k=1}^{(i-1)}$ , and  $\{\mathbf{T}_k^{(n-1)}\}_{k=i+1}^p$  fixed, Theorem 1 computes the optimal transform  $\mathbf{T}_i^{(n)}$  to minimize  $D_{MSE, \mathbf{W}}$ . A similar argument shows that the error term cannot increase. The distortion sequence  $\{D_{MSE, \mathbf{W}}(n)\}$  is nonincreasing and nonnegative; hence  $\lim_{n \rightarrow \infty} D_{MSE, \mathbf{W}}(n) = \inf\{D_{MSE, \mathbf{W}}(n)\}$  by monotone convergence. ■

*Remark 4:* The local convergence in Theorem 2 is affected by several factors: (i) The covariance structure  $\Sigma_{\mathbf{x}}$  of the source; (ii) The DAG structure of  $G$ ; (iii) The schedule of iterative optimization of local matrices



and factorization of  $\mathbf{T}$  into the  $\mathbf{T}_i$ ; (iv) The random initialization of  $\{\mathbf{T}_i\}_{i=1}^p$ . In practice, multiple executions of Algorithm 1 increase the probability of converging to a global minimum.

#### F. Example: A Multi-Hop Network

Consider the noiseless multi-hop network of Fig. 3 in which a relay aggregates, compresses and/or forwards its observations to a receiver. The network is a hybrid combination of a distributed and point-to-point network.

*Example 4 (“Hybrid Network”):* High-dimensional, correlated signals  $\mathbf{x}_1 \in \mathbb{R}^{n_1}$  and  $\mathbf{x}_2 \in \mathbb{R}^{n_2}$  are observed at nodes  $v_1$  and  $v_2$  where  $n_1 = n_2 = 15$  dimensions. The covariance  $\Sigma_{\mathbf{x}}$  of the global source  $\mathbf{x} = [\mathbf{x}_1; \mathbf{x}_2]$  was generated as follows for the experiment, ensuring  $\Sigma_{\mathbf{x}} \succ \mathbf{0}$ . The diagonal entries  $(i, i)$  of  $\Sigma_{\mathbf{x}}$  were selected as  $15 + 2U_{ii}$ , and off-diagonal entries  $(i, j)$  for  $j > i$  were selected as  $1 + 2U_{ij}$  where  $U_{ii}$  and  $U_{ij}$  are *i.i.d.* uniform random variables over the interval  $[0, 1]$ .

The linear transfer function is factored in the form  $\mathbf{T} = \mathbf{T}_1\mathbf{T}_2$  where  $\mathbf{T}_1 = \mathbf{L}_{34}$  and

$$\mathbf{T}_2 = \begin{bmatrix} \mathbf{L}_{13} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{23} \end{bmatrix}.$$

The target reconstruction at  $v_4$  is the entire signal  $\mathbf{r}_4 = \mathbf{x}$ . The bandwidth  $c_{34} = 11$ , while bandwidth  $c = c_{13} = c_{23}$  is varied for the experiment. Depending on the amount of bandwidth  $c$ , the network operates in one of the modes given in Table I. Fig. 3(b) plots the sum distortion vs. compression performance, and Fig. 3(c) plots the convergence of Algorithm 1 for the operating point  $c = 6$ ,  $c_{34} = 11$ .

### V. NOISY NETWORKS

We now analyze communication for networks with non-ideal channels:  $\mathbf{y}_{ij} = \mathbf{x}_{ij} + \mathbf{z}_{ij}$ . Edges  $(i, j)$  represent vector Gaussian channels. Network communication is limited according to both bandwidth compression ratios  $\alpha_{ij}$  and signal-to-noise ratios  $SNR_{ij}$ . We simplify optimization of subspaces by restricting attention to single-layer multi-source, multi-receiver networks for which  $\mathcal{V} = \mathcal{S} \cup \mathcal{T}$ . In this case, the linear transfer function is  $\mathbf{y} = \mathbf{T}\mathbf{x} + \mathbf{z}$ , i.e. the noise is additive but not filtered over multiple network layers.

#### A. MSE Distortion at Receivers

Each receiver  $v_i \in \mathcal{T}$  receives observations  $\mathbf{y}_i = \mathbf{G}_i\mathbf{x} + \mathbf{z}_i$  where  $\mathbf{z}_i$  is the noise to  $v_i$ . The MSE distortion for reconstructing  $\mathbf{r}_i$  at receiver  $v_i$  is given by,

$$\begin{aligned} \tilde{D}_i &= \text{tr}(\Sigma_{\mathbf{r}}) - 2\text{tr}(\mathbf{B}_i\mathbf{G}_i\Sigma_{\mathbf{x}}\mathbf{r}_i) + \text{tr}(\mathbf{B}_i\Sigma_{\mathbf{z}_i}\mathbf{B}_i^T) \\ &\quad + \text{tr}(\mathbf{B}_i\mathbf{G}_i\Sigma_{\mathbf{x}}\mathbf{G}_i^T\mathbf{B}_i^T). \end{aligned} \tag{35}$$

Setting the matrix derivative with respect to  $\mathbf{B}_i$  in Eqn. (35) to zero yields the optimal linear transform  $\mathbf{B}_i$  (cf. Eqn. (18)),

$$\mathbf{B}_i^{opt} = \Sigma_{\mathbf{r}_i \mathbf{x}} \mathbf{G}_i^T \left[ \mathbf{G}_i \Sigma_{\mathbf{x}} \mathbf{G}_i^T + \Sigma_{\mathbf{z}_i} \right]^{-1}. \quad (36)$$

Combining the LLSE estimates as  $\hat{\mathbf{r}} = \mathbf{B}\mathbf{y}$ , where  $\mathbf{y} = \mathbf{T}\mathbf{x} + \mathbf{z}$ , the weighted MSE for all receivers is given by

$$\begin{aligned} \tilde{D}_{MSE, \mathbf{W}} &= E \left[ \|\mathbf{r} - \hat{\mathbf{r}}\|_{\mathbf{W}}^2 \right] \\ &= E \left[ \|\mathbf{r} - \mathbf{B}(\mathbf{T}\mathbf{x} + \mathbf{z})\|_{\mathbf{W}}^2 \right] \\ &= \text{tr}(\mathbf{W}\mathbf{B}\mathbf{T}\Sigma_{\mathbf{x}}\mathbf{T}^T\mathbf{B}^T\mathbf{W}^T) - 2\text{tr}(\mathbf{W}\mathbf{B}\mathbf{T}\Sigma_{\mathbf{x}\mathbf{r}}\mathbf{W}^T) \\ &\quad + \text{tr}(\mathbf{W}\Sigma_{\mathbf{r}}\mathbf{W}^T) + \text{tr}(\mathbf{W}\mathbf{B}\Sigma_{\mathbf{z}}\mathbf{B}^T\mathbf{W}^T). \end{aligned} \quad (37)$$

By construction of the weighting matrix  $\mathbf{W}$ , the MSE in Eqn. (37) is a weighted sum of individual distortions at receivers, i.e.  $\tilde{D}_{MSE, \mathbf{W}} = \sum_i w_i \tilde{D}_i$ .

### B. Computing Encoding Transform $\mathbf{T}$

For noisy networks, power constraints on channel inputs limit the amount of amplification of transmitted signals. For single-layer networks, let  $v_i \in \mathcal{S}$  be a source node with observed signal  $\mathbf{x}_i$ . A power constraint on the input to channel  $(i, j) \in \mathcal{E}$  is given by

$$E[\|\mathbf{x}_{ij}\|_2^2] = E[\|\mathbf{L}_{ij}\mathbf{x}_i\|_2^2] = \text{tr}(\mathbf{L}_{ij}\Sigma_{\mathbf{x}_i}\mathbf{L}_{ij}^T) \leq P_{ij}. \quad (38)$$

The power constraint in Eqn. (38) is a quadratic function of the entries of the global linear transform  $\mathbf{T}$ . More precisely, let  $\boldsymbol{\ell}_{ij} = \text{vec}(\mathbf{L}_{ij})$  and  $\mathbf{t} = \text{vec}(\mathbf{T})$ . Since  $\mathbf{t}$  contains all variables of  $\boldsymbol{\ell}_{ij}$ , we may write  $\boldsymbol{\ell}_{ij} = \mathbf{J}_{ij}\mathbf{t}$  where  $\mathbf{J}_{ij}$  selects variables from  $\mathbf{t}$ . Using the matrix-vector identities of Eqn. (21), the power constraint in Eqn. (38) can be written as

$$\begin{aligned} \text{tr}(\mathbf{L}_{ij}\Sigma_{\mathbf{x}_i}\mathbf{L}_{ij}^T) &= \boldsymbol{\ell}_{ij}^T (\Sigma_{\mathbf{x}_i} \otimes \mathbf{I}) \boldsymbol{\ell}_{ij} \\ &= \mathbf{t}^T \mathbf{J}_{ij}^T (\Sigma_{\mathbf{x}_i} \otimes \mathbf{I}) \mathbf{J}_{ij} \mathbf{t}. \end{aligned} \quad (39)$$

Letting  $\boldsymbol{\Gamma}_{ij} \triangleq \mathbf{J}_{ij}^T (\Sigma_{\mathbf{x}_i} \otimes \mathbf{I}) \mathbf{J}_{ij}$ , the quadratic constraint is  $\mathbf{t}^T \boldsymbol{\Gamma}_{ij} \mathbf{t} \leq P_{ij}$ . The matrix  $\boldsymbol{\Gamma}_{ij}$  is a symmetric, positive semi-definite matrix. Thus a power constraint is a quadratic, convex constraint.

**Algorithm 2** NOISY-COMPRESSION-ESTIMATION( $\mathcal{N}, \mathbf{W}, \epsilon$ )

- 
- 1: Identify compression matrix  $\mathbf{T}$  and corresponding linear equality constraints  $(\Phi, \phi)$ , and quadratic power constraints  $\{(\Gamma_{ij}, P_{ij})\}_{(i,j) \in \mathcal{E}}$ . Identify estimation matrices  $\{\mathbf{B}_i\}_{i:v_i \in \mathcal{T}}$ . [Sec. III, Sec. V-B]
  - 2: Initialize  $\mathbf{T}^{(0)}$  randomly to a feasible matrix.
  - 3: Set  $n = 1$ ,  $\tilde{D}_{MSE, \mathbf{W}}(0) = \infty$ .
  - 4: **repeat**
  - 5:   Compute  $\{\mathbf{B}_i^{(n)}\}_{i:v_i \in \mathcal{T}}$  given  $\mathbf{T}^{(n-1)}$ . [Eqn. (36)]
  - 6:   Compute  $\mathbf{T}^{(n)}$  given  $\{\mathbf{B}_i^{(n)}\}_{i:v_i \in \mathcal{T}}$ ,  $(\Phi, \phi)$ ,  $\{(\Gamma_{ij}, P_{ij})\}_{(i,j) \in \mathcal{E}}$ . [Theorem 3]
  - 7:   Compute  $\tilde{D}_{MSE, \mathbf{W}}(n)$ . [Eqn. (37)]
  - 8:   Set  $\tilde{\Delta}_{MSE, \mathbf{W}} = \tilde{D}_{MSE, \mathbf{W}}(n) - \tilde{D}_{MSE, \mathbf{W}}(n-1)$ .
  - 9:   Set  $n = n + 1$ .
  - 10: **until**  $\tilde{\Delta}_{MSE, \mathbf{W}} \leq \epsilon$  or  $n \geq N_{max}$ .
  - 11: **return**  $\mathbf{T}^{(n)}$  and  $\{\mathbf{B}_i^{(n)}\}_{i:v_i \in \mathcal{T}}$ .
- 

*C. Quadratic Program with Convex Constraints*

As in Section IV-B, we use the vector form  $\mathbf{t} = \text{vec}(\mathbf{T})$  to enforce linear equality constraints  $\Phi \mathbf{t} = \phi$ . For noisy networks, we include power constraints  $\mathbf{t}^T \Gamma_{ij} \mathbf{t} \leq P_{ij}$  for each channel  $(i, j) \in \mathcal{E}$ . For a fixed global decoding transform  $\mathbf{B}$ , the distortion  $\tilde{D}_{MSE, \mathbf{W}}$  of Eqn. (37) is again a quadratic function of  $\mathbf{t}$ . Using the compact notation

$$p \triangleq \text{tr}(\mathbf{W} \Sigma_{\mathbf{r}} \mathbf{W}^T) + \text{tr}(\mathbf{W} \mathbf{B} \Sigma_{\mathbf{z}} \mathbf{B}^T \mathbf{W}^T), \quad (40)$$

$$\mathbf{p} \triangleq -2 \text{vec}(\mathbf{B}^T \mathbf{W}^T \mathbf{W} \Sigma_{\mathbf{r} \mathbf{x}}), \quad (41)$$

$$\mathbf{P} \triangleq \Sigma_{\mathbf{x}} \otimes \mathbf{B}^T \mathbf{W}^T \mathbf{W} \mathbf{B}, \quad (42)$$

a derivation identical to that of Lemma 1 yields  $\tilde{D}_{MSE, \mathbf{W}} = \mathbf{t}^T \mathbf{P} \mathbf{t} + \mathbf{p}^T \mathbf{t} + p$ . The optimal encoding transform  $\mathbf{T}$  for single-layer noisy networks is solvable via a quadratic program with quadratic constraints (QCQP), following the development of Eqns. (40)-(42), and the power constraints given in Eqns. (38)-(39); cf. Theorem 1.

*Theorem 3 (Optimal Encoding  $\mathbf{T}$  for Noisy LTN):* Let  $\mathcal{N}$  be a single-layer LTN,  $\mathbf{B}$  be the fixed decoding transform, and  $\mathbf{t} = \text{vec}(\mathbf{T})$  be the encoding transform. The optimal encoding  $\mathbf{t}$  is the solution to the following

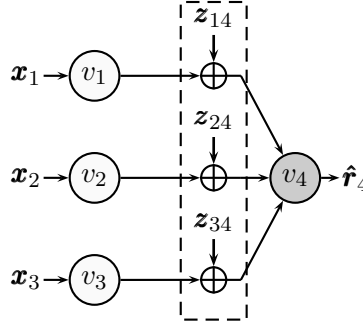


Fig. 4. A block diagram of a distributed, noise/power limited LTN. Each source node transmits signal projections of a vector  $\mathbf{x}_i \in \mathbb{R}^4$  to a decoder over a vector arbitrary white Gaussian noise (AWGN) channel.

quadratic program with quadratic constraints (QCQP):

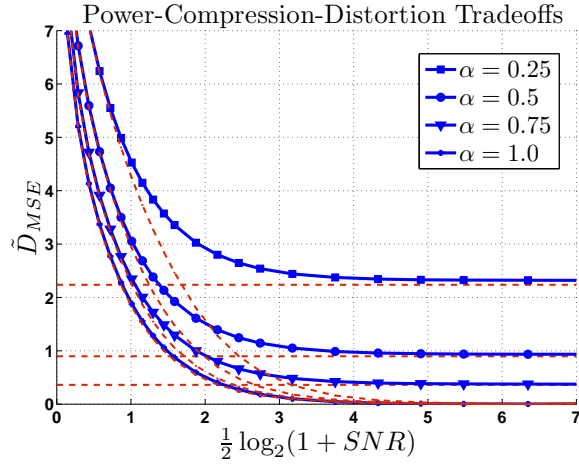
$$\begin{aligned} \arg \min_{\mathbf{t}} \quad & \mathbf{t}^T \mathbf{P} \mathbf{t} + \mathbf{p}^T \mathbf{t} + p \\ \text{s. t.} \quad & \mathbf{\Phi} \mathbf{t} = \phi, \\ & \mathbf{t}^T \mathbf{\Gamma}_{ij} \mathbf{t} \leq P_{ij}, \quad (i, j) \in \mathcal{E}, \end{aligned} \tag{43}$$

where  $(\mathbf{\Phi}, \phi)$  represent linear equality constraints (dictated by network topology), and  $\{(\mathbf{\Gamma}_{ij}, P_{ij})\}_{(i,j) \in \mathcal{E}}$  represent quadratic power constraints on variables of  $\mathbf{T}$ .

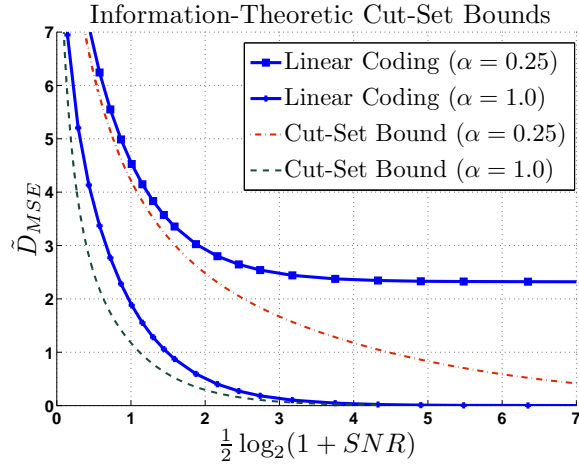
*Remark 5:* A quadratic program with linear and convex quadratic constraints is solvable efficiently via standard convex program solvers; the time complexity depends polynomially on the number of matrix variables and constraints.

#### D. Iterative Algorithm and Convergence

Algorithm 2 defines an iterative algorithm for single-layer, noise/power limited networks. In addition to subspace selection, the amount of power per subspace is determined iteratively. The iterative method alternates between optimizing the global decoding transform  $\mathbf{B}$  and the global encoding transform  $\mathbf{T}$ , ensuring that network topology and power constraints are satisfied. As in Theorem 2, the weighted MSE distortion is a nonincreasing function of the iteration number, i.e.  $\tilde{D}_{MSE, \mathbf{W}}(n) \geq \tilde{D}_{MSE, \mathbf{W}}(n+1)$ . While convergence to a stationary point is guaranteed, the optimization space is highly complex— a globally optimal solution is not guaranteed.



(a) COMPRESSION-ESTIMATION TRADEOFFS



(b) CUT-SET LOWER BOUNDS (INFORMATION THEORY)

Fig. 5. (a) Power-compression-distortion “spectra” of the network for varying compression ratios  $\alpha$  and  $SNR$  levels. The (red, unmarked) dashed lines represent cut-set lower bounds on achievable MSE distortions for linear coding based on convex relaxations discussed in Section VI-E. (b) For  $\alpha \in \{0.25, 1.0\}$ , the performance of linear coding is compared with information-theoretic cut-set bounds (described in Section VI-F). In the high-SNR setting, information-theoretic coding strategies are capable of zero-distortion; however, in the low-SNR setting, linear coding achieves a competitive MSE performance while maintaining zero-delay and low-complexity.

#### E. Example: A Distributed Noisy Network

Fig. 4 diagrams a classic example of a distributed network with multiple source (sensor) nodes transmitting signal projections to a central decoder. Each source node is power constrained and must transmit a compressed description of its observed signal over a noisy vector channel.

*Example 5 (Distributed LTN):* In Fig. 4, the global source  $\mathbf{x} = [\mathbf{x}_1; \mathbf{x}_2; \mathbf{x}_3]$  is chosen to be a *jointly Gaussian* vector with  $n = 12$  dimensions, and  $n_i = 4$  for each of  $|\mathcal{S}| = 3$  source nodes. Here, we specify the exact distribution of  $\mathbf{x}$  in order to provide information-theoretic lower bounds. We set the covariance of  $\mathbf{x}$  to be Gauss-Markov with  $\rho = 0.8$ ,

$$\Sigma_{\mathbf{x}} = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{11} \\ \rho & 1 & \rho & \dots & \rho^{10} \\ \rho^2 & \rho & 1 & \dots & \rho^9 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{11} & \rho^{10} & \rho^9 & \dots & 1 \end{bmatrix}.$$

The network structure is specified by bandwidths  $c_{14} = c_{24} = c_{34} = c$ . The global encoding transform  $\mathbf{T}$  is block-diagonal with matrices  $\mathbf{L}_{14}$ ,  $\mathbf{L}_{24}$ , and  $\mathbf{L}_{34}$  on the diagonal. The compression ratio is varied equally for each source node,  $\alpha = \frac{c}{n_i}$  where  $n_i = 4$ . The noise variables  $\mathbf{z}_{ij}$  are *i.i.d.* Gaussian random vectors with zero-mean and identity covariances. The power constraints are set as  $P_1 = P_2 = P_3 = c(SNR)$ , where  $SNR_{ij} = SNR$  for all links. The goal of destination  $v_4$  is to reconstruct the entire source  $\mathbf{r}_4 = \mathbf{x}$ . Fig. 5(a) plots the performance of LTN optimization for varying  $\alpha$  and  $SNR$  ratios as well as cut-set lower bounds for linear coding based on convex relaxations. Cut-set lower bounds for linear coding for this example are explained further in Section VI-E. Fig. 5(b) plots cut-set bounds based on information theory which are explained further in Sections VI-F and VI-G.

*Remark 6 (Comparison with [5], [6]):* For this example, as the  $SNR \rightarrow \infty$ , the error  $\tilde{D}_{MSE}$  approaches the error associated to the distributed KLT [5] where channel noise was not considered. In [6], the authors model the effects of channel noise; however, they do not provide cut-set lower bounds. In addition, the iterative optimization of the present paper optimizes all compression matrices simultaneously per iteration and allows arbitrary convex constraints, as opposed to the schemes in both [5], [6] which optimize the encoding matrix of each user separately per iteration.

## VI. CUT-SET LOWER BOUNDS

In this section, we derive lower bounds on the minimum MSE distortion possible for linear compression and estimation of correlated signals in the LTN model. Our main technique is to relax an arbitrary acyclic graph along all possible graph cuts to point-to-point networks with side information. The cut-set bounds provide a performance benchmark for the iterative methods of Sections IV-V.

### A. Point-to-Point Network with Side Information

Consider the point-to-point network of Fig. 6. Source node  $v_1$  compresses source  $\mathbf{x} \in \mathbb{R}^n$  via a linear transform  $\mathbf{L}_{12}$ . The signal  $\mathbf{x}_{12} \in \mathbb{R}^{c_{12}}$  is transmitted where  $\mathbf{x}_{12} = \mathbf{L}_{12}\mathbf{x}$  and  $E[\|\mathbf{x}_{12}\|_2^2] \leq P$ . Receiver  $v_2$  computes a linear estimate of desired signal  $\mathbf{r} \in \mathbb{R}^r$  using observations  $\mathbf{y}_{12} = \mathbf{x}_{12} + \mathbf{z}$  and side information  $\mathbf{s} \in \mathbb{R}^s$  as follows,

$$\hat{\mathbf{r}} = \mathbf{B} \begin{bmatrix} \mathbf{y}_{12} \\ \mathbf{s} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{12} \\ \mathbf{s} \end{bmatrix}. \quad (44)$$

The decoding transform  $\mathbf{B}$  is here partitioned into two sub-matrices  $\mathbf{B}_{11}$  and  $\mathbf{B}_{12}$ . We will find it convenient to define the following random vectors,

$$\boldsymbol{\xi} \triangleq \mathbf{x} - \boldsymbol{\Sigma}_{\mathbf{x}\mathbf{s}}\boldsymbol{\Sigma}_{\mathbf{s}}^{-1}\mathbf{s}, \quad (45)$$

$$\boldsymbol{\nu} \triangleq \mathbf{r} - \boldsymbol{\Sigma}_{\mathbf{r}\mathbf{s}}\boldsymbol{\Sigma}_{\mathbf{s}}^{-1}\mathbf{s}. \quad (46)$$

Signals  $\boldsymbol{\xi}$  and  $\boldsymbol{\nu}$  are innovation vectors. For example,  $\boldsymbol{\xi}$  is the difference between  $\mathbf{x}$  and the linear least squares estimate of  $\mathbf{x}$  given  $\mathbf{s}$  which is equivalent to  $\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{s}}\boldsymbol{\Sigma}_{\mathbf{s}}^{-1}\mathbf{s}$ .

### B. Case I: Ideal Vector Channel

In the ideal case,  $P = \infty$  or  $\mathbf{z} = 0$ . The weighted, linear minimum MSE distortion of the point-to-point network with side information is obtained by solving

$$\begin{aligned} D_{ideal}^* &= \min_{\mathbf{L}_{12}, \mathbf{B}} E[\|\mathbf{r} - \hat{\mathbf{r}}\|_{\mathbf{W}}^2], \\ &= \min_{\mathbf{L}_{12}, \mathbf{B}_{11}, \mathbf{B}_{12}} E[\|\mathbf{r} - (\mathbf{B}_{11}\mathbf{L}_{12}\mathbf{x} + \mathbf{B}_{12}\mathbf{s})\|_{\mathbf{W}}^2]. \end{aligned} \quad (47)$$

The following theorem specifies the solution to Eqn. (47).

*Theorem 4 (Ideal Network Relaxation):* Let  $\mathbf{x} \in \mathbb{R}^n$ ,  $\mathbf{s} \in \mathbb{R}^s$ , and  $\mathbf{r} \in \mathbb{R}^r$  be zero-mean random vectors with given full-rank covariance matrices  $\boldsymbol{\Sigma}_{\mathbf{x}}$ ,  $\boldsymbol{\Sigma}_{\mathbf{s}}$ ,  $\boldsymbol{\Sigma}_{\mathbf{r}}$  and cross-covariances  $\boldsymbol{\Sigma}_{\mathbf{r}\mathbf{x}}$ ,  $\boldsymbol{\Sigma}_{\mathbf{r}\mathbf{s}}$ ,  $\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{s}}$ . Let  $\boldsymbol{\xi}$  and  $\boldsymbol{\nu}$  be the innovations defined in Eqn (45) and Eqn. (46) respectively. The solution to the minimization of Eqn. (47) over matrices  $\mathbf{L}_{12} \in \mathbb{R}^{c_{12} \times n}$ ,  $\mathbf{B}_{11} \in \mathbb{R}^{r \times c_{12}}$ , and  $\mathbf{B}_{12} \in \mathbb{R}^{r \times s}$  is obtained in closed form as

$$D_{ideal}^* = \text{tr}(\boldsymbol{\Sigma}_{\boldsymbol{\nu}}\mathbf{W}^T\mathbf{W}) - \sum_{j=1}^{c_{12}} \lambda_j, \quad (48)$$

where  $\{\lambda_j\}_{j=1}^{c_{12}}$  are the  $c_{12}$  largest eigenvalues of the matrix  $\mathbf{W}\boldsymbol{\Sigma}_{\boldsymbol{\nu}\boldsymbol{\xi}}\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\nu}}\mathbf{W}^T$ .

*Proof:* The optimization in Eqn. (47) is simplified by first determining the LMMSE optimal  $\mathbf{B}_{12}$  transform in terms of  $\mathbf{B}_{11}$  and  $\mathbf{L}_{12}$ :  $\mathbf{B}_{12}^{opt} = \boldsymbol{\Sigma}_{\mathbf{r}\mathbf{s}}\boldsymbol{\Sigma}_{\mathbf{s}}^{-1} - \mathbf{B}_{11}\mathbf{L}_{12}\boldsymbol{\Sigma}_{\mathbf{x}\mathbf{s}}\boldsymbol{\Sigma}_{\mathbf{s}}^{-1}$ . Plugging  $\mathbf{B}_{12}^{opt}$  into Eqn. (47)

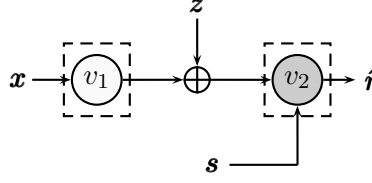


Fig. 6. A point-to-point network with side information  $\mathbf{s}$  at the receiver. In the case of additive noise  $\mathbf{z}$ , the input to the channel is power-constrained so that  $E[\|\mathbf{x}_{12}\|_2^2] \leq P$ .

yields a minimization over  $\mathbf{B}_{11}$  and  $\mathbf{L}_{12}$  only. By grouping and rearranging variables in terms of innovation vectors  $\boldsymbol{\xi}$  and  $\boldsymbol{\nu}$ ,

$$D_{ideal}^* = \min_{\mathbf{L}_{12}, \mathbf{B}_{11}} E[\|\boldsymbol{\nu} - \mathbf{B}_{11}\mathbf{L}_{12}\boldsymbol{\xi}\|_{\mathbf{W}}^2]. \quad (49)$$

The optimization of Eqn. (49) is that of an equivalent point-to-point network with input signal  $\boldsymbol{\xi}$  and desired reconstruction  $\boldsymbol{\nu}$ , without side information. Eqn. (49) is in standard form and solvable using canonical correlation analysis as detailed in [34, p. 368]. The optimal value  $D_{ideal}^*$  is given in Eqn. (48) in terms of the eigenvalues of  $\mathbf{W}\boldsymbol{\Sigma}_{\boldsymbol{\nu}\boldsymbol{\xi}}\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\nu}}\mathbf{W}^T$ . ■

### C. Case II: Additive Noise and Power Constraints

In the case of additive noise  $\mathbf{z}$  (here with assumed covariance  $\boldsymbol{\Sigma}_{\mathbf{z}} = \mathbf{I}$  for compactness) and a power-constrained input to the vector channel, the weighted, linear minimum MSE distortion is obtained by solving

$$\begin{aligned} D_{noisy}^* &= \min_{\mathbf{L}_{12}, \mathbf{B}_{11}, \mathbf{B}_{12}} E[\|\mathbf{r} - (\mathbf{B}_{11}(\mathbf{L}_{12}\mathbf{x} + \mathbf{z}) + \mathbf{B}_{12}\mathbf{s})\|_{\mathbf{W}}^2], \\ \text{s.t. } &\text{tr}[\mathbf{L}_{12}\boldsymbol{\Sigma}_{\mathbf{x}}\mathbf{L}_{12}^T] \leq P. \end{aligned} \quad (50)$$

Again, by solving for the optimal LMMSE matrix  $\mathbf{B}_{12}$  and grouping terms in the resulting optimization according to innovation vectors  $\boldsymbol{\xi}$  and  $\boldsymbol{\nu}$ ,

$$\begin{aligned} D_{noisy}^* &= \min_{\mathbf{L}_{12}, \mathbf{B}_{11}} E[\|\boldsymbol{\nu} - (\mathbf{B}_{11}(\mathbf{L}_{12}\boldsymbol{\xi} + \mathbf{z}))\|_{\mathbf{W}}^2], \\ \text{s.t. } &\text{tr}[\mathbf{L}_{12}\boldsymbol{\Sigma}_{\mathbf{x}}\mathbf{L}_{12}^T] \leq P. \end{aligned} \quad (51)$$

*Remark 7:* The exact solution to Eqn. (51) involves handling a quadratic power constraint and a rank constraint due to the reduced-dimensionality of  $\mathbf{L}_{12}$ . In [6, Theorem 4], a related optimization problem was solved via a Lagrangian relaxation. For our problem, we take a simpler approach using a semi-definite



programming (SDP) relaxation. We first note that  $D_{noisy}^* \geq D_{ideal}^*$ . In the high-SNR regime, the two distortion values are asymptotically equivalent. Therefore, we compute a good approximation for the distortion  $D_{noisy}^*$  in the low-SNR regime via the following SDP relaxation.

*Theorem 5 (SDP Relaxation):* Consider random vectors  $\mathbf{x}, \mathbf{s}, \mathbf{r}, \boldsymbol{\xi}, \boldsymbol{\nu}$ , and matrices  $\mathbf{L}_{12}, \mathbf{B}_{11}$  as defined in Theorem 4. In addition, let random vector  $\mathbf{z}$  have zero-mean and covariance  $\boldsymbol{\Sigma}_{\mathbf{z}} = \mathbf{I}$ . Let  $\boldsymbol{\Psi} \triangleq \mathbf{L}_{12}^T \mathbf{L}_{12}$  and  $\boldsymbol{\Phi} \in \mathbb{R}^{r \times r}$  be an arbitrary positive semi-definite matrix where  $r$  is the dimension of random vector  $\mathbf{r}$ . The following lower bound applies,

$$\begin{aligned} D_{noisy}^* \geq \min_{\boldsymbol{\Phi}, \boldsymbol{\Psi}} \quad & \text{tr}[\boldsymbol{\Phi}] + \text{tr}[\mathbf{W}[\boldsymbol{\Sigma}_{\boldsymbol{\nu}} - \boldsymbol{\Sigma}_{\boldsymbol{\nu}\boldsymbol{\xi}}\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\nu}}]\mathbf{W}^T], \\ \text{s.t.} \quad & \text{tr}[\boldsymbol{\Sigma}_{\mathbf{x}}\boldsymbol{\Psi}] \leq P, \quad \boldsymbol{\Psi} \succeq \mathbf{0}, \\ & \begin{bmatrix} \boldsymbol{\Phi} & \mathbf{W}\boldsymbol{\Sigma}_{\boldsymbol{\nu}\boldsymbol{\xi}}\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1} \\ \boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\xi}\boldsymbol{\nu}}\mathbf{W}^T & \boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1} + \boldsymbol{\Psi} \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \quad (52)$$

The proof of Theorem 5 is based on a rank relaxation as detailed in the Appendix. The power constraint is still enforced in Eqn. (52). In the low-SNR regime, power allocation over subspaces dominates the error performance. If we denote the solution to the SDP of Theorem 5 as  $D_{sdp}^*$ , we arrive at the following characterization,

$$D_{noisy}^* \geq \max\{D_{ideal}^*, D_{sdp}^*\}. \quad (53)$$

#### D. Cut-Set Lower Bounds for Linear Coding

Consider an LTN graph  $\mathcal{N}$  with source nodes  $\mathcal{S} \subset \mathcal{V}$  and receivers  $\mathcal{T} \subset \mathcal{V}$ . We assume that  $\mathcal{S} \cap \mathcal{T} = \emptyset$ , i.e. the set of sources and receivers are disjoint. The total bandwidth and total power across a *cut*  $\mathcal{F} \subset \mathcal{V}$  are defined respectively as

$$C(\mathcal{F}) = \sum_{\substack{jk \in \mathcal{E} \\ j \in \mathcal{F}, k \in \mathcal{F}^c}} c_{jk}, \quad (54)$$

$$P(\mathcal{F}) = \sum_{\substack{jk \in \mathcal{E} \\ j \in \mathcal{F}, k \in \mathcal{F}^c}} P_{jk}, \quad (55)$$

where the edge set  $\mathcal{E}$  and bandwidths  $c_{jk}$  were defined in Section II. The edges of the graph are directed, hence the bandwidth across a cut accounts for the  $c_{ij}$  only for those edges directed from node  $v_i$  to  $v_j$ . In the following theorem, the notation  $\mathbf{x}_{\mathcal{F}}$  denotes the concatenation of vectors  $\mathbf{x}_i : v_i \in \mathcal{F}$ . The set  $\mathcal{F}^c$  denotes the complement of  $\mathcal{F}$  in  $\mathcal{V}$ .

*Definition 10:*  $D_{ideal}^*[\mathbf{x}, \mathbf{r} | \mathbf{s}; c, \mathbf{W}]$  represents the distortion  $D_{ideal}^*$  computed with the weighted norm via  $\mathbf{W}$  for the ideal point-to-point network with input  $\mathbf{x}$ , bandwidth  $c$ , reconstruction vector  $\mathbf{r}$ , and side information to receiver  $\mathbf{s}$ . Similarly,  $D_{noisy}^*[\mathbf{x}, \mathbf{r} | \mathbf{s}; c, P, \mathbf{W}]$  represents the distortion  $D_{noisy}^*$  for a noisy point-to-point network with channel-input power constraint  $P$  and noise vector  $\mathbf{z}$  with zero-mean with identity covariance.

*Theorem 6 (Cut-Set Lower Bounds):* Let  $\mathcal{N}$  be an arbitrary LTN graph with source nodes  $\mathcal{S}$  and receivers  $\mathcal{T}$ . Let  $\mathcal{F} \subset \mathcal{V}$  be a cut of the graph. For ideal channel communication,

$$E\left[\|\mathbf{r}_{\mathcal{F}^c} - \hat{\mathbf{r}}_{\mathcal{F}^c}\|_{\mathbf{W}}^2\right] \geq D_{ideal}^*[\mathbf{x}_{\mathcal{F}}, \mathbf{r}_{\mathcal{F}^c} | \mathbf{x}_{\mathcal{F}^c}; C(\mathcal{F}), \mathbf{W}]. \quad (56)$$

In the case of noisy channel communication over network  $\mathcal{N}$  with additive channel noise  $\mathbf{z}_{ij}$  (assumed zero-mean, identity covariance),

$$\begin{aligned} E\left[\|\mathbf{r}_{\mathcal{F}^c} - \hat{\mathbf{r}}_{\mathcal{F}^c}\|_{\mathbf{W}}^2\right] \\ \geq D_{noisy}^*[\mathbf{x}_{\mathcal{F}}, \mathbf{r}_{\mathcal{F}^c} | \mathbf{x}_{\mathcal{F}^c}; C(\mathcal{F}), P(\mathcal{F}), \mathbf{W}]. \end{aligned} \quad (57)$$

*Proof:* The LTN graph is partitioned into two sets  $\mathcal{F}$  and  $\mathcal{F}^c$ . The source nodes  $v_i \in \mathcal{F}$  are merged as one source “super” node, and the receivers  $v_i \in \mathcal{F}^c$  are merged into one receiver “super” node. The maximum bandwidth and maximum power between the source and receiver are  $C(\mathcal{F})$  and  $P(\mathcal{F})$  respectively. The random vector  $\mathbf{x}_{\mathcal{F}^c}$  represents those signals with channels to the receiver super node, not accounted for in the cut  $\mathcal{F}$ ; hence, this information is given as side information (a relaxation) to the receiver. The relaxed network after the merging process is the point-to-point network of Fig. 6 with noise  $\mathbf{z}$  of dimension equal to the bandwidth  $C(\mathcal{F})$  of the cut, and provides a lower bound on the MSE distortion  $E\left[\|\mathbf{r}_{\mathcal{F}^c} - \hat{\mathbf{r}}_{\mathcal{F}^c}\|_{\mathbf{W}}^2\right]$  at receivers  $v_i \in \mathcal{F}^c$ . ■

*Remark 8:* The total number of distinct cuts  $\mathcal{F}$  separating sources and receivers is  $(2^{|\mathcal{S}|} - 1)(2^{|\mathcal{T}|} - 1)$ . For a particular cut, there exists a continuum of lower bounds for multi-receiver networks depending on the choice of weighting  $\mathbf{W}$ .

#### E. Example: Cut-Set Lower Bounds for Linear Coding

In Fig. 5(a), cut-set lower bounds for linear coding are illustrated based on Theorem 6 for a distributed noisy network. The bounds are depicted for the cut that separates all sources from the receiver. Due to our approximation method in Eqn. (53) based on the SDP relaxation, the lower bounds show tight agreement in the low-SNR and high-SNR asymptotic regimes.

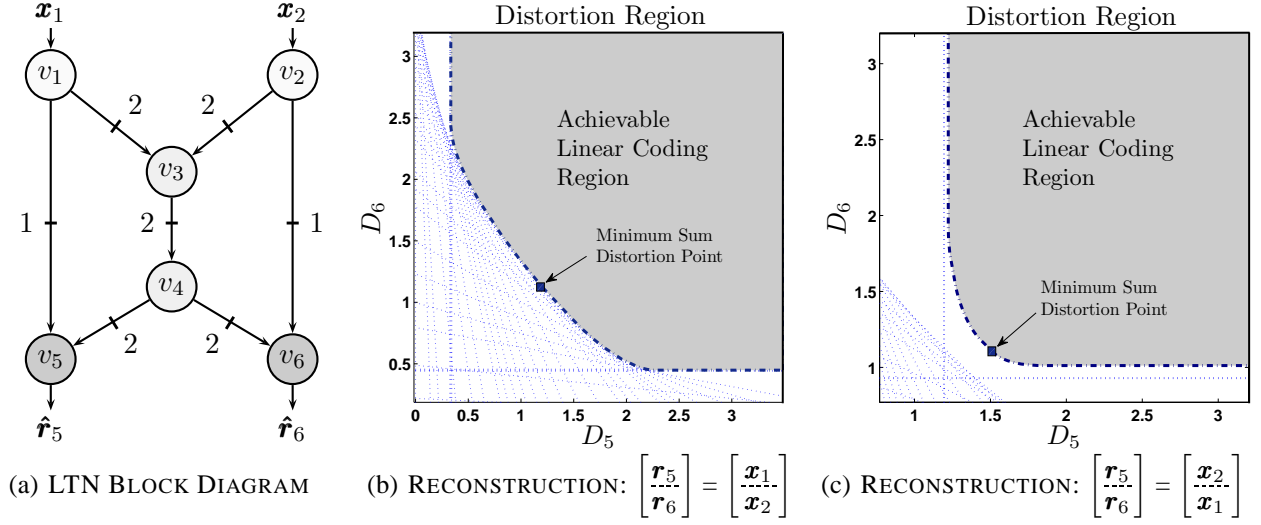


Fig. 7. (a) Block diagram of a multi-source, multi-destination ideal network. Bandwidths  $c_{ij}$  of all links are labeled. Although the graph is symmetric, the source covariance matrix given in Eqn. (63) includes cross-correlations which cause the distortion plots to appear asymmetric. (b) The distortion region assuming that node  $v_5$  reconstructs  $\mathbf{x}_1$ , and node  $v_6$  reconstructs  $\mathbf{x}_2$ . The cut-set lower bounds are drawn as dotted lines, and the shaded region depicts the achievable points. (c) The distortion region assuming that node  $v_5$  reconstructs  $\mathbf{x}_2$ , and node  $v_6$  reconstructs  $\mathbf{x}_1$ .

#### F. Cut-Set Lower Bound From Information Theory

For the point-to-point communication scenario illustrated in Fig. 6, the information-theoretically optimal performance can be determined precisely. Consider an  $\ell$ -length sequence  $\{(\mathbf{x}[t], \mathbf{s}[t])\}_{t=1}^{\ell}$  of jointly *i.i.d.* random vectors. The source node  $v_1$  has access to the source sequence  $\{\mathbf{x}[t]\}_{t=1}^{\ell}$ . We will assume throughout that  $\mathbf{r}$  (respectively  $\mathbf{r}[t]$ ) is a deterministic function of  $(\mathbf{x}, \mathbf{s})$  (respectively  $(\mathbf{x}[t], \mathbf{s}[t])$ ). The goal of receiver  $v_2$  is to minimize the *average* MSE distortion  $D_{\ell} = E \left[ \frac{1}{\ell} \sum_{t=1}^{\ell} \|\mathbf{r}[t] - \hat{\mathbf{r}}[t]\|_2^2 \right]$  where the reconstruction sequence  $\{\hat{\mathbf{r}}[t]\}_{t=1}^{\ell}$  is generated based on access to side information  $\{\mathbf{s}[t]\}_{t=1}^{\ell}$  and the sequence of channel output vectors. We study the performance in the limit as  $\ell \rightarrow \infty$  and denote  $D \triangleq D_{\infty}$ .

1) *Source-Channel Separation*: We establish a lower bound by combining the data processing inequality with the definitions of Wyner-Ziv rate-distortion function and channel capacity. Specifically, by straightforward extension of [35], the minimum rate  $R(D)$  required to reconstruct  $\{\mathbf{r}[t]\}_{t=1}^{\infty}$  at distortion  $D$  is given by  $R(D) = \min I(\mathbf{x}; \mathbf{u} | \mathbf{s})$  where the minimization is over all “auxiliary” random vectors  $\mathbf{u}$  for which  $p(\mathbf{u}, \mathbf{x}, \mathbf{s}) = p(\mathbf{u} | \mathbf{x}) p(\mathbf{x}, \mathbf{s})$  and for which  $E[\|\mathbf{r} - E[\mathbf{r} | \mathbf{u}, \mathbf{s}]\|_2^2] \leq D$ . Furthermore, by definition of the channel

capacity  $C(P)$  between  $v_1$  and  $v_2$ ,  $C(P) = \max_{p(\mathbf{x}_{12}): E[\|\mathbf{x}_{12}\|_2^2] \leq P} I(\mathbf{x}_{12}; \mathbf{y}_{12})$ .<sup>3</sup> Source-channel separation applies to the scenario of Fig. 6, and in a nearly identical proof as detailed in [36, Thm. 1.10],

$$R(D) \leq C(P). \quad (58)$$

2) *R(D) for Jointly Gaussian Sources:* If  $\{\mathbf{r}[t], \mathbf{x}[t], \mathbf{s}[t]\}$  form an *i.i.d.* sequence of jointly Gaussian random vectors, then  $R(D)$  is equal to the conditional rate-distortion function [5, Appendix II],

$$R_c(D) = \min_{p(\hat{\mathbf{r}}|\mathbf{x}, \mathbf{s}): E[\|\mathbf{r} - \hat{\mathbf{r}}\|_2^2] \leq D} I(\mathbf{x}; \hat{\mathbf{r}}|\mathbf{s}). \quad (59)$$

3) *Capacity of the Vector AWGN Channel:* If the channel noise  $\mathbf{z}$  is a Gaussian random vector with zero mean and covariance  $\Sigma_{\mathbf{z}} = \mathbf{I}$ , the capacity of the channel in Fig. 6 with bandwidth  $c_{12}$  and power constraint  $P$  is

$$C(P) = \frac{c_{12}}{2} \log_2 \left[ 1 + \frac{P}{c_{12}} \right]. \quad (60)$$

4) *Cut-set Bound:* We utilize Eqn. (58) to obtain an information-theoretic lower bound to the distortion achievable in any network of the type considered in this paper. An arbitrary graph is reduced via graph cuts to point-to-point networks. The following theorem collects the known information-theoretic results discussed.

*Theorem 7 (Cut-Set Bounds: Info. Theory):* Let  $\mathcal{N}$  be an arbitrary LTN graph with vector AWGN channels. Consider a cut  $\mathcal{F} \subset \mathcal{V}$  separating the graph into a point-to-point network with bandwidth  $C(\mathcal{F})$  and power  $P(\mathcal{F})$ . Let  $R(D_{opt}^*)$  be the rate-distortion function for the source  $\mathbf{x}_{\mathcal{F}}$  with side information  $\mathbf{x}_{\mathcal{F}^c}$  and reconstruction  $\mathbf{r}_{\mathcal{F}^c}$ .<sup>4</sup> Then

$$R(D_{opt}^*) \leq \frac{C(\mathcal{F})}{2} \log_2 \left[ 1 + \frac{P(\mathcal{F})}{C(\mathcal{F})} \right]. \quad (61)$$

#### G. Example: Cut-Set Lower Bound From Information Theory

For the noisy network in Example 5, consider cut  $\mathcal{F} = \{v_1, v_2, v_3\}$ . The source signal  $\mathbf{x}_{\mathcal{F}} = \mathbf{x} = [\mathbf{x}_1; \mathbf{x}_2; \mathbf{x}_3]$  is jointly Gaussian, the side information is absent, and  $\mathbf{r}_{\mathcal{F}^c} = \mathbf{x}$ . Denote the eigenvalues of the source  $\mathbf{x}_{\mathcal{F}}$  as  $\{\lambda_{\mathbf{x}, i}\}_{i=1}^n$ . Evaluating Eqn. (59) as in [5, Appendix II], optimal source coding corresponds to reverse water-filling over the eigenvalues (see also [37, Chap. 10]),

<sup>3</sup>The notation in information theory vs. signal processing differs. The term  $I(\mathbf{x}_{12}; \mathbf{y}_{12})$  denotes the mutual information between random vectors whereas the term  $p(\mathbf{x}_{12})$  indicates a probability distribution.

<sup>4</sup>We assume that  $\mathbf{r}_{\mathcal{F}^c}$  is a deterministic function of the global source  $\mathbf{x}$ .

$$R_c(D_{opt}^*) = \sum_{i=1}^n \max \left\{ \frac{1}{2} \log_2 \frac{\lambda_{\mathbf{x},i}}{D_i}, 0 \right\},$$

$$\text{where } D_i = \begin{cases} \theta & \text{if } \theta < \lambda_{\mathbf{x},i} \\ \lambda_{\mathbf{x},i} & \text{if } \theta \geq \lambda_{\mathbf{x},i} \end{cases}$$

and where  $\theta$  is chosen such that  $\sum_{i=1}^n D_i = D_{opt}^*$ . The lower bound of Eqn. (61) is plotted in Fig. 5(b) for two different bandwidth compression ratios.

#### H. Example: Multi-Source, Multi-Receiver Network

*Example 6 (Multiple Unicast):* In Fig. 7, the global source  $\mathbf{x} = [\mathbf{x}_1; \mathbf{x}_2]$  where  $\mathbf{x}_1 \in \mathbb{R}^4$  and  $\mathbf{x}_2 \in \mathbb{R}^4$ . The correlation structure of  $\mathbf{x}$  is given by the following matrices,

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} 2.4 & 1.1 & 0.4 & 0.0 & 0.1 & 0.1 & 0.0 & 0.1 \\ 1.1 & 1.7 & 0.8 & 0.4 & 0.0 & 0.2 & 0.2 & 0.1 \\ 0.4 & 0.8 & 1.2 & 0.0 & 0.2 & 0.6 & 0.1 & 0.3 \\ 0.0 & 0.4 & 0.0 & 0.8 & 0.3 & 0.0 & 0.1 & 0.0 \\ 0.1 & 0.0 & 0.2 & 0.3 & 1.1 & 0.1 & 0.2 & 0.0 \\ 0.1 & 0.2 & 0.6 & 0.0 & 0.1 & 1.2 & 0.2 & 0.1 \\ 0.0 & 0.2 & 0.1 & 0.1 & 0.2 & 0.2 & 1.0 & 0.6 \\ 0.1 & 0.1 & 0.3 & 0.0 & 0.0 & 0.1 & 0.6 & 1.2 \end{bmatrix}. \quad (63)$$

The network structure is specified by bandwidths  $c_{ij}$  as labeled in Fig. 7(a). The factorization of the global linear transform  $\mathbf{T}$  was given in Example 2 of Section IV.

The distortion region for the network in the case when node  $v_5$  estimates  $\mathbf{r}_5 = \mathbf{x}_1$ , and node  $v_6$  estimates  $\mathbf{r}_6 = \mathbf{x}_2$  is given in Fig. 7(b). A direct link exists from each source to receiver. However, if the desired reconstruction at the receivers is switched as in Fig. 7(c), the channel from  $v_3$  to  $v_4$  must be shared fully and becomes a bottleneck. The cut-set bounds of interest are shown in dotted lines. The shaded region depicts the points achievable via the iterative method of Section IV. In Fig. 7(c), the upper and lower bounds are not tight everywhere—even if one receiver is completely ignored, the resulting problem is still a distributed compression problem for which tight bounds are not known. The achievable curve was generated by taking the convex hull of 32 points corresponding to weighting ratios  $\frac{w_5}{w_6} \in [\frac{1}{100}, 100]$ .

In Table II, we compare the results of linear transform design methods for the minimum sum distortion point (weighting ratio  $\frac{w_5}{w_6} = 1$ ).

- *Random Projections*— Each entry for all compression matrices is selected from the standard normal distribution. The sum distortion  $D_5 + D_6$  is averaged over  $10^2$  random compression matrices selected for all nodes.
- *Routing and Network Coding (Ad-Hoc)*— For the scenario in Fig. 7(b), nodes  $v_1$  and  $v_2$  project their signal onto the principal eigenvectors of  $\Sigma_{11}$  and  $\Sigma_{22}$  respectively. Routing permits each receiver to

TABLE II  
COMPARISON OF REDUCED-DIMENSION LINEAR TRANSFORMS

<b>Design Method</b>	<i>Fig. 5(b)</i>	<i>Fig. 5(c)</i>
	$D_5 + D_6$	$D_5 + D_6$
<i>Random Projections</i>	4.3170	6.3471
<i>Routing and Network Coding</i>	2.7029	3.8170
<i>Iterative QP Optimization</i>	2.3258	2.6165
<i>⟨Lower Bound⟩</i>	2.3243	2.3243

receive the best two eigenvector projections from its corresponding source, as well as an extra projection from the other source. For Fig. 7(c), using a simple “network coding” strategy of adding signals at  $v_3$ , one receiver is able to receive its best two eigenvector projections, but the other receiver can only receive one best eigenvector projection.

- *Iterative QP Optimization*– Linear transforms are designed using the iterative method of Section IV.
- *Lower Bound*– The minimum sum distortion possible due to the cut-set lower bound of Theorem 6.

## VII. CONCLUSION

The linear transform network (LTN) was proposed to model the aggregation, compression, and estimation of correlated random signals in directed, acyclic graphs. For both noiseless and noisy LTN graphs, a new iterative algorithm was introduced for the joint optimization of reduced-dimension network matrices. Cut-set lower bounds were introduced for zero-delay linear coding based on convex relaxations. Cut-set lower bounds for optimal coding were introduced based on information-theoretic principles. The compression-estimation tradeoffs were analyzed for several example networks. A future challenge remains to compute tighter lower bounds and relaxations for non-convex network optimization problems. Reduced-dimension linear transforms have potential applications in data fusion and sensor networks. The idea of exploiting correlations between network signals to reduce data transmission, and the idea of approximate reconstruction as opposed to exact recovery at receivers may lead to further advances in networking.

## APPENDIX

Starting from the optimization in Eqn. (51), the LLSE optimal matrix  $\mathbf{B}_{11}^{opt} = \Sigma_{\nu\xi}\mathbf{L}_{12}^T(\mathbf{L}_{12}\Sigma_{\xi}\mathbf{L}_{12}^T + \mathbf{I})^{-1}$ , assuming  $\Sigma_{\mathbf{z}} = \mathbf{I}$ . Substituting this expression and simplifying the objective function in Eqn. (51),

$$\begin{aligned} D_{noisy}^* = \min_{\mathbf{L}_{12}} \quad & \text{tr} [\mathbf{W}\Sigma_{\nu}\mathbf{W}^T] \\ & + \text{tr} \left[ \mathbf{W}\Sigma_{\nu\xi}\mathbf{L}_{12}^T [\mathbf{L}_{12}\Sigma_{\xi}\mathbf{L}_{12}^T + \mathbf{I}]^{-1} \mathbf{L}_{12}\Sigma_{\xi\nu}\mathbf{W}^T \right] \\ \text{s.t.} \quad & \text{tr} [\mathbf{L}_{12}\Sigma_{\mathbf{x}}\mathbf{L}_{12}^T] \leq P. \end{aligned} \quad (64)$$

Applying the Woodbury (matrix-inversion) identity [33, C.4.3] to the objective function and simplifying terms,

$$\begin{aligned} D_{noisy}^* = \min_{\mathbf{L}_{12}} \quad & \text{tr} [\mathbf{W}\Sigma_{\nu}\mathbf{W}^T] - \text{tr} \left[ \mathbf{W}\Sigma_{\nu\xi}\Sigma_{\xi}^{-1}\Sigma_{\xi\nu}\mathbf{W}^T \right] \\ & + \text{tr} \left[ \mathbf{W}\Sigma_{\nu\xi}\Sigma_{\xi}^{-1} \left[ \Sigma_{\xi}^{-1} + \mathbf{L}_{12}^T\mathbf{L}_{12} \right]^{-1} \Sigma_{\xi}^{-1}\Sigma_{\xi\nu}\mathbf{W}^T \right] \\ \text{s.t.} \quad & \text{tr} [\mathbf{L}_{12}\Sigma_{\mathbf{x}}\mathbf{L}_{12}^T] \leq P. \end{aligned} \quad (65)$$

Introducing a positive semi-definite matrix  $\Phi$  such that  $\Phi \succeq \mathbf{W}\Sigma_{\nu\xi}\Sigma_{\xi}^{-1}[\Sigma_{\xi}^{-1} + \mathbf{L}_{12}^T\mathbf{L}_{12}]^{-1}\Sigma_{\xi}^{-1}\Sigma_{\xi\nu}\mathbf{W}^T$ , written equivalently in Schur-complement form [33, A.5.5], and setting  $\Psi = \mathbf{L}_{12}^T\mathbf{L}_{12} \in \mathbb{R}^{n \times n}$  as a rank  $c_{12}$  matrix,

$$\begin{aligned} D_{noisy}^* = \min_{\Phi, \Psi} \quad & \text{tr} [\Phi] + \text{tr} \left[ \mathbf{W}[\Sigma_{\nu} - \Sigma_{\nu\xi}\Sigma_{\xi}^{-1}\Sigma_{\xi\nu}]\mathbf{W}^T \right], \\ \text{s.t.} \quad & \text{tr} [\Sigma_{\mathbf{x}}\Psi] \leq P, \quad \Psi \succeq \mathbf{0}, \quad \text{rank} [\Psi] = c_{12}, \\ & \begin{bmatrix} \Phi & \mathbf{W}\Sigma_{\nu\xi}\Sigma_{\xi}^{-1} \\ \Sigma_{\xi}^{-1}\Sigma_{\xi\nu}\mathbf{W}^T & \Sigma_{\xi}^{-1} + \Psi \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \quad (66)$$

Dropping the rank constraint yields the relaxation of Eqn. (52).

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

## REFERENCES

- [1] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, April 2006.
- [2] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [3] V. Goyal, "Theoretical foundations of transform coding," *IEEE Signal Processing Magazine*, vol. 18, pp. 9–21, Sept. 2001.

- [4] A. d'Aspremont, L. E. Ghaoui, M. I. Jordan, and G. R. Lanckriet, "A direct formulation for sparse pca using semidefinite programming," *SIAM Review*, vol. 49, no. 3, pp. 434–448, 2007.
- [5] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed karhunen loève transform," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5177–5196, 2006.
- [6] I. D. Schizas, G. B. Giannakis, and Z. Q. Luo, "Distributed estimation using reduced-dimensionality sensor observations," *IEEE Transactions on Signal Processing*, vol. 55, pp. 4284–4299, Aug. 2007.
- [7] J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 757–770, 2008.
- [8] K. Zhang, *Best Linear Unbiased Estimation Fusion with Constraints*. PhD thesis, Univ. of New Orleans, New Orleans, LA, Dec. 2003.
- [9] J. Fang and H. Li, "Optimal/near-optimal dimensionality reduction for distributed estimation in homogeneous and certain inhomogeneous scenarios," *IEEE Transactions on Signal Processing*, vol. 58, pp. 4339–4353, Aug. 2010.
- [10] Y. Zhu, E. Song, J. Zhou, and Z. You, "Optimal dimensionality reduction of sensor data in multisensor estimation fusion," *IEEE Transactions on Signal Processing*, vol. 53, pp. 1631–1639, May 2005.
- [11] J. Fang and H. Li, "Joint dimension assignment and compression for distributed multisensor estimation," *IEEE Signal Processing Letters*, vol. 15, pp. 174–177, Jan. 2008.
- [12] T. Ramstad, "Shannon mappings for robust communication," *Teletronikk*, vol. 98, no. 1, pp. 114–128, 2002.
- [13] M. Skoglund, N. Phamdo, and F. Alajaji, "Hybrid digital-analog source-channel coding for bandwidth compression/expansion," *IEEE Transactions on Information Theory*, vol. 52, pp. 3757–3763, Aug. 2006.
- [14] F. Hekland, P. A. Floor, and T. A. Ramstad, "Shannon-kotel'nikov mappings in joint source-channel coding," *IEEE Transactions on Communications*, vol. 57, pp. 94–105, Jan. 2009.
- [15] A. Ramamoorthy, K. Jain, P. A. Chou, and M. Effros, "Separating distributed source coding from network coding," *IEEE Transactions on Information Theory*, vol. 52, pp. 2785–2795, June 2006.
- [16] S. Jalali and M. Effros, "On the separation of lossy source-network coding and channel coding in wireline networks," in *Proc. IEEE International Symposium on Information Theory*, pp. 500–504, June 2010.
- [17] M. Gastpar, "On source-channel communication in networks," in *Advances in Network Information Theory* (P. Gupta, G. Kramer, and A. van Wijngarden, eds.), DIMACS, pp. 217–237, March 2003.
- [18] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed karhunen-loève transform," in *Proc. IEEE Workshop on Multimedia Signal Processing*, pp. 57–60, Dec. 2002.
- [19] H. I. Nurdin, R. R. Mazumdar, and A. Bagchi, "Reduced-dimension linear transform coding of distributed correlated signals with incomplete observations," *IEEE Transactions on Information Theory*, vol. 55, pp. 2848–2858, June 2009.
- [20] O. Roy and M. Vetterli, "Dimensionality reduction for distributed estimation in the infinite dimensional regime," *IEEE Transactions on Information Theory*, vol. 54, pp. 1655–1669, April 2008.
- [21] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, pp. 1204–1216, July 2000.
- [22] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, pp. 371–381, Feb. 2003.
- [23] S. Jaggi, P. Sanders, P. Chou, M. Effros, S. Egner, K. Jain, and L. Tolhuizen, "Polynomial time algorithms for multicast network code construction," *IEEE Transactions on Information Theory*, vol. 51, pp. 1973–1982, June 2005.
- [24] L. Ford and D. Fulkerson, "Maximal flow through a network," *Canadian Journal of Mathematics*, vol. 8, pp. 399–404, 1956.



- [25] R. Koetter and F. R. Kschischang, "Coding for errors and erasures in random network coding," *IEEE Transactions on Information Theory*, vol. 54, pp. 3579–3591, Aug. 2008.
- [26] D. Silva, F. R. Kschischang, and R. Koetter, "A rank-metric approach to error control in random network coding," *IEEE Transactions on Information Theory*, vol. 54, pp. 3951–3967, Sept. 2008.
- [27] A. Wiesel and A. O. Hero III, "Decomposable principal component analysis," *IEEE Transactions on Signal Processing*, vol. 57, pp. 4369–4377, Nov. 2009.
- [28] G. Shen, S. K. Narang, and A. Ortega, "Adaptive distributed transforms for irregularly sampled wireless sensor networks," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, April 2009.
- [29] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Joint source-channel communication for distributed estimation in sensor networks," *IEEE Transactions on Information Theory*, vol. 53, pp. 3629–3653, Oct. 2007.
- [30] M. Rabbat, J. Haupt, A. Singh, and R. Nowak, "Decentralized compression and predistribution via randomized gossiping," in *Proc. Int. Conf. on Information Processing in Sensor Networks*, IPSN, 2006.
- [31] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," 2005. Available at: <http://dsp.rice.edu/cs>.
- [32] P. Healy and N. S. Nikolov, "How to layer a directed acyclic graph," in *Revised Papers from the 9th International Symposium on Graph Drawing*, Graph Drawing, pp. 16–30, 2002.
- [33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge: Cambridge University Press, 2004.
- [34] D. R. Brillinger, *Time Series: Data Analysis and Theory*. San Francisco: Holden-Day, 1981.
- [35] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, pp. 1–10, January 1976.
- [36] M. Gastpar, *To Code or Not To Code*. PhD thesis, EPFL, 2002.
- [37] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.